BRL-TR-3089

4

DTIC FILE COPY

TECHNICAL REPORT BRL-TR-3089

BRL

AD-A220 150

EXTREME QUANTILE ESTIMATION IN BINARY RESPONSE MODELS

BARRY A. BODT
HENRY B. TINGEY

MARCH 1990

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

U.S. ARMY LABORATORY COMMAND

BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

90 04 06 053

NOTICES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
March 1990

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Extreme Quantile Estimation in Binary Response Models

5. FUNDING NUMBERS

6. AUTHOR(S)
Barry A. Bodt and Henry B. Tingey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Ballistic Research Laboratory
ATTN: SLCBR-DD-T
Aberdeen Proving Ground, MD 21005-5066

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
BRL-TR-3089

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Binary response models are used in estimating the probability, P(x), of a successful response when an experimental unit is exposed to a stimulus level, x. Frequently, the aim of the investigator is to locate a specific level of stimulus for which P(x) takes on a known value, p. The stimulus level sought, X_{100p}, is referred to as the P^{th} quantile of P(x). When p lies outside the interval [.25,.75], X_{100p} is considered to be an extreme quantile. The focus of this paper is the estimation of extreme X_{100p}.

All estimation techniques for extreme quantiles seek to limit estimate bias, lack of precision, and the importance of parametric assumption. The extreme quantile estimate described here utilizes, through a transformation of the responses, median estimation techniques where such problems are not so acute. The feasibility of the approach is demonstrated in a Monte Carlo study. It is shown that the design aspect of the approach has properties important to c-optimality, and it is argued that most data sets should be fit well by the model employed. This paper also explores the relationship between data structure and the existence of maximum likelihood estimates in finite samples.

14. SUBJECT TERMS
Sequential Process
Experimental Design
Sensitivity Testing
C-optimality
Optimal Design

15. NUMBER OF PAGES
88

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only (*Leave blank*).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- DOE - See authorities.
- NASA - See Handbook NHB 2200.2.
- NTIS - Leave blank.

Block 12b. Distribution Code.

- DOD - Leave blank.
- DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- NASA - Leave blank.
- NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Professor John W. Green of the University of Delaware and Dr. Malcolm S. Taylor of the Ballistic Research Laboratory for their technical suggestions. Also, the authors thank Mrs. Annette L. Wiseman for her contributions to the typing of this paper and Mr. Vernon V. Visnaw and Mrs. Linda L.C. Moss for their efforts in reviewing this work.

INTENTIONALLY LEFT BLANK.

TABLE OF CONTENTS

TABLE OF CONTENTS (contd)

LIST OF ILLUSTRATIONS

INTENTIONALLY LEFT BLANK.

LIST OF TABLES

INTENTIONALLY LEFT BLANK.

1. INTRODUCTION

1.1 Background

Binary response models are used in estimating the performance sensitivity of a subject population exposed to levels of a stimulus. The model arises as follows. Assume that the stimulus influences performance and the problem is only to describe the nature of this influence. For an individual subject, performance can be classified as either a response or nonresponse, where a response to the stimulus is viewed as a successful performance. It is assumed that a response occurs only when the applied stimulus exceeds the subject's unknown tolerance, the stimulus level above which the subject is sensitive. When characterizing the population, we denote performance in terms of the probability of observing a response for each level of stimulus, that is, the true proportion of the population with tolerances less than that level. This probability corresponds to the distribution function of subject tolerance. The binary response model imposes a problem structure through which performance sensitivity can be expressed in terms of an estimated tolerance distribution.

Binary response models have two basic applications: to allow experimenters to choose among several populations according to which has the more desirable sensitivity; or alternatively, to allow experimenters to seek a stimulus to which the population is more sensitive, or a specific level of stimulus for which an acceptable number of responses are likely to be observed. A few examples demonstrate the widespread applicability of these models.

Ballisticians test the performance of a penetrator by firing it against a target and assessing the damage, where damage is defined as perforation or nonperforation of the target. The resulting damage relates directly to the penetrator's striking velocity. A response curve characterizes this relationship by indicating the probability of a response (perforation) for each fixed level of velocity. In effect the response curve conveys, in a probabilistic sense, how sensitive to velocity is the performance of the penetrator population. A penetrator deemed insensitive over a standard velocity range is considered undesirable for use as a threat mechanism.

Other examples lending themselves to sensitivity analysis include the determination of the quantity of poison necessary to kill a rodent, the tensile strength required to withstand a stress, or the armor thickness needed to repel a bullet. The analysis in these cases might suggest a need for recommended levels of dosage, stress, or armor thickness, or, alternatively, improvements in poison potency, tensile strength, or armor material. The common structure of these problems is made apparent in the next section.

1.2 Statistical Problem Statement

In this section we define the structure of the modeling problem, present the general approach that is used, and explain why the approach is reasonable. We begin by describing the data. The data in a sensitivity test environment are characterized by three common features. First, it is assumed that the stimulus does affect subject performance. Second, when a stimulus is applied to a subject the result is one of two possible outcomes, response or nonresponse. Third, a subject cannot be exposed to more than one stimulus level because the subject properties change with their first stimulus exposure. Restated, the second and third conditions describe a Bernoulli trial in which the testing is destructive.

The principal goal of the analysis of sensitivity test data is the estimation of the response curve P(x) for all or some levels, x, of the stimulus. So far, no restriction has been placed on the model P(x), but we know from Section 1.1 that the purpose of the model is to convey information about the performance of the subject population for various levels of the stimulus. Expressing performance in terms of the proportion of favorable results is a natural approach, and from the random selection of subjects this proportion may be viewed as a probability. Thus, the first restriction is that P(x) must be a probability for each stimulus level. Still, in terms of modeling we have only restricted the range to [0,1]. We now impose further constraints. Let us assume the real-valued response curve has the following properties:

1. $P(-\infty) = 0,$ (1.1)

2. $P(\infty) = 1,$

3. P(x) is strictly increasing,

4. P(x) is continuous.

These restrictions imply the response curve P(x) is a distribution function, but certainly a stimulus-response relationship need not assume such a form. For example, consider the performance of a drug in its ability to cure an illness. If no drug is administered the patient may still regain his health; thus, a probability of zero may never be encountered. If excessive amounts of the drug are used, at some dosage detrimental effects may result which would contradict the monotonicity property. But for a variety of applications the conditions imposed are not constraining. For the ballistics example, a zero velocity will obviously cause a failure to perforate, and an infinite velocity will definitely cause a perforation to occur. Assuming an infinite population, the physics of the test suggest further that the continuity and monotonicity properties would not be unexpected. When these model limitations are acceptable, it is convenient to think of P(x) as being the distribution of a specific random variable.

The random variable tolerance arises as follows. We assume that each member of the subject population has a tolerance to the stimulus. For a specific subject, application of any stimulus above its tolerance necessarily results in a response. Application of a stimulus at or below the tolerance results in a nonresponse. Assuming a continuum for the mapping, tolerance is a continuous random variable which is not directly observable. Rather its value can only be bounded through the observance of a response or nonresponse, e.g., a stimulus causing a nonresponse must be less than or equal to the tolerance of the subject. The realization of tolerance for a subject is that subject's specific sensitivity to the stimulus variable.

To summarize, sensitivity analysis using the binary response model expresses the relationship between some stimulus variable and the resultant probability of response for the subject population. Assuming that this response curve adheres to the conditions of a distribution function, one can conceive of a random variable (tolerance) with physical significance which would have that exact distribution. The response curve is then identical to the tolerance distribution, and estimation of some interval or quantile of this distribution becomes the task.

1.3. Purpose

Complete knowledge of the tolerance or response distribution provides precise information regarding the subject population's sensitivity to the stimulus variable. A general p^{th} quantile (x_{100p}) yields less complete knowledge but often contains sufficient information for valuable inference; specifically, it represents the stimulus at which 100p percent of the subject population responds. The standard measure in many sensitivity environments is x_{50}. Trevan [1927] first suggested the use of the median dose in the context of biological assay. Today the median effective dose (ED_{50}) and the median lethal dose (LD_{50}) serve as baselines for comparisons among drugs. For example, in sensitivity analyses where drug selection is the goal, response distributions are often assumed to be similar, that is, differing only in location. Thus, differences in performance could be determined by comparing the estimates of x_{100p} for any general p; however, the median dose is usually used. One advantage to using x_{50} is that the asymptotic variance of $\hat{x}_{100p}$ achieves a minimum at $p=.5$ for the common methods used. Moreover, several Monte Carlo studies involving these methods support the minimum variance property for small samples as well. See, for example, Wetherill [1963].

Some studies require information about the subject population's sensitivity for which the x_{50} is not well suited. For example, it is of limited practical value to know the armor thickness which will permit perforation by fifty percent of the threat mechanisms. Quantiles in the tail of the response distribution contain more useful information in this context. The utility of extreme quantiles in practice was recognized by C.I. Bliss as quoted from Brown [1967]:

> ... interest does not always center on the ED_{50} or LD_{50}. Sometimes an extreme percentage is important. For example, in sterilization tests for fruit flies the quarantine officials desired 0 percent survival. It took some arguments to convince them that it is impossible to measure 0 or 100 percent. Another example that arises in therapeutics is determination of the 'safety margin', that is, the difference between curative and lethal doses. Here interest might center on estimating the ED_{99} (the dose that cures 99 percent) and the LD_{01} (the dose that kills one percent). Actually, the ED_{95} and the LD_{05} are preferable for realistic points.

Extreme quantiles, though useful for inference, remain difficult to estimate in many practical settings. "Some methods are provided for estimating more general points on a response curve ... , but extreme percentage points should be avoided" [Wetherill 1963]. Although progress has been made since Wetherill's 1963 paper, the issue is as yet unresolved.

In this paper we will develop an alternative to the current procedures for the estimation of extreme quantiles.

1.4 Estimation Procedures for x_{100p}

Approaches to the estimation of x_{100p} are varied. Methods include the use of both fixed and sequential designs for data collection; for either, selection of the stimulus levels may incorporate nonparametric or parametric considerations. Many parametric assumptions are in use and include the normal, logistic, and Weibull distributions. Several different estimators are often appropriate for use under the same design. Hamilton [1979] compared the performance of ten different estimators for x_{50}, all drawing upon the same data. Hybrid strategies combine methods usually treated separately; for example, data gathered from a nonparametric sequential design may be used in forming a parametric maximum likelihood estimate. In the following, estimation procedure refers to any design and estimation combination.

The variety of possible procedures has stimulated much research. An extensive review of the literature is not given here. Instead we give a brief summary for general x_{100p} in Sections 1.4.1-1.4.3 with special focus on those procedures and results germane to our specific interest--extreme quantile estimation. The attention given to x_{50} is necessary as background for later development. A detailed review of literature targeting extreme quantiles appears in Chapter 2.

1.4.1 Estimation Techniques

Well-known nonparametric techniques include the work of Kärber [1931], Wetherill et al. [1966], and Robbins and Monro [1951]. The Spearman-Kärber method and Wetherill's $\bar{w}$ estimate only the x_{50}, while the Stochastic Approximation

Method of Robbins and Monro [1951] estimates any general x_{100p}. Generally, these nonparametric estimators focus attention on a specific quantile with no formal way of estimating neighboring quantiles with known accuracy and precision. We are interested in estimating quantiles neighboring the design's "target quantile" as well. This is done easily if the response distribution form is known. Therefore, we choose to approach estimation parametrically.

When a parametric assumption can be made for the response distribution, minimum chi-square and maximum likelihood estimation are commonly used. For the application intended here, the number of observations taken from a given distributional class is small. A practical disadvantage of minimum chi-square estimation is that these limited samples may cause numerical instability, driven by very small expected frequencies for some classes [Finney 1978]. In maximum likelihood estimation, small class frequency is not as serious a problem. As to their relative performance within the context of sensitivity experiments,

> ... no clear ruling can be given that one method is generally better than the other in its approach to the true values of the parameters for either normal or logistic models, and indeed it seems unlikely that a consistent superiority of either will ever be demonstrated [Finney 1978].

Considering this position we will concentrate on maximum likelihood estimation.

Maximum likelihood estimates (MLEs) of location and scale are easily developed for the traditionally used two-parameter distributions. Under the usual parameterization of the response distribution, $P(x) = F(\alpha + \eta x)$ for a completely specified $F(\cdot)$, where α and η are the location and scale parameters respectively. The likelihood function is given by

$$L = \prod_{i=1}^{k} \binom{n_i}{r_i} P(x_i)^{r_i} (1-P(x_i))^{n_i - r_i},$$

where r_i/n_i is the observed proportion of responses for stimulus x_i. Solution of the following equations yields MLEs for α and η. Denoting l as the log-likelihood, we have

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{k} \left[\frac{r_i - n_i P(x_i)}{P(x_i)\,(1 - P(x_i))} \right] \frac{\partial F}{\partial \alpha} (\alpha + \eta x_i) = 0 \qquad (1.2)$$

$$\frac{\partial l}{\partial \eta} = \sum_{i=1}^{k} \left[\frac{r_i - n_i P(x_i)}{P(x_i)(1 - P(x_i))} \right] x_i \frac{\partial F}{\partial \eta} (\alpha + \eta x_i) = 0$$

By the parameterization $P(x) = F(\alpha + \eta x)$, it follows that

$$x_{100p} = \frac{\gamma_{100p} - \alpha}{\eta}, \tag{1.3}$$

where γ_{100p} is the p^{th} quantile of $F(\cdot)$. Then by the invariance property of maximum likelihood, the MLE of x_{100p} is given by the right side of (1.3), with $\hat{\alpha}$ and $\hat{\eta}$ substituted for the true parameters. The well known efficiency and consistency properties of MLEs may be used to develop asymptotic results. We defer this development to Chapter 3 where a specific parametric form is considered.

Maximum likelihood estimation is possible with any parametric form for which appropriate regularity conditions hold. Among those are probit, normit, logit, linit, and more recently, quantit transformations as well as several non "it" forms. The first two correspond to the normal distribution, and the second two refer to the logistic and uniform distributions respectively. These and other historical parametric forms are discussed in Finney [1978]. The quantit transformation is based on a three parameter distribution given by Mielke [1972] in the context of rank tests. It was suggested for use in sensitivity analysis by Copenhaver and Mielke [1977]. It is representative of recent efforts by Einbinder [1973], Prentice [1976], Aranda-Urdaz [1981], and Guerrero and Johnson [1982] to generalize the parametric form assumed for the response distribution. All involve more than two parameters and are considered generalizations because, for each, special cases result in common response distribution forms such as the normal and logistic. When the response distribution form is not known, a more general parametric model lends greater credence to the resulting estimates [Prentice 1976].

1.4.2 Design

An experimental design determines the levels of stimulus to be considered and the number of subjects to be tested at each level. Both fixed and sequential designs are used in sensitivity testing. Which approach is preferred depends on many factors including the experimenter's knowledge of the response distribution, the number of available subjects, the quantile of primary interest, the time allotted for testing, and the practicable range of the stimulus. For estimation of extreme quantiles, the majority of the literature suggests the implementation of a sequential design.

Some notable exceptions to this rule are the fixed designs of Chernoff [1962], Little [1976], and Hoel and Jennrich [1979]. Chernoff determined designs which minimize, assuming a normal response distribution, the asymptotic variance of $\hat{x}_{100p}$. They are discussed in more detail in Section 2.2.2. Little suggested allocating samples according to linear regression techniques. From $P(x) = F(\alpha + \eta x)$ consider that $F^{-1}(P(x))$ is a linear function in x. His strategy consists of allocating samples to two stimulus levels corresponding to moderately high and low probabilities, respectively, in proportions so as to minimize the variance of the extrapolated extreme quantile estimate. He developed the designs for the normal, logistic, and extreme value distributions. Hoel and Jennrich also approached the problem from a regression standpoint. They used an optimal extrapolation design for a Chebyshev regression model to allocate samples for the estimation of lower extreme quantiles. The response distribution form for which this was done is given by

$$P(x) = 1 - e^{-\sum_{j=0}^{k} \alpha_j x^j} .$$

Since all of these optimal designs were based on a parametric assumption, their applicability should depend, at least partially, on that assumption. They also addressed the robustness issue. They considered two situations where only the family of P(x) was chosen correctly. Altering the coefficients had little effect on the selection of an optimal design for the two cases examined.

Returning to sequential strategies, one advantage is their ability to reliably allocate more samples in the region of interest, that is, near the quantile to be estimated. We say more samples because sequential procedures generally tend to converge to the region of interest if not to the quantile itself. This ability need not be tied to a restrictive parametric assumption, and the importance of this property in relation to optimal designs is discussed in Chapter 3. We introduce a sequential procedure in Chapter 2 based on one of the designs which follow.

The best known designs are the work of Dixon and Mood [1948] and Robbins and Monro [1951]. Neither were intended for use in extreme quantile estimation. Their relevance to this task is made apparent in Chapter 2. Dixon and Mood [1948] described the Up and Down method in reference to finding the median tolerance for a population of explosives. The procedure calls for the prior selection of "potential" levels of stimulus which cover the entire stimulus range and are spaced with a common, fixed distance between levels. After the selection of an initial design point, sampling proceeds (one subject at a time) by moving up one level if a nonresponse is observed at the current level and moving down if a response is observed. With the simplest interpretation of $P(x)$ (the probability of observing a response) and reasonable spacing between levels, the tendency of this design to sample about the median is intuitive.

The Stochastic Approximation Method of Robbins and Monro [1951] locates the quantile x_{100p} by finding a solution to $P(x) - p = 0$. We emphasize that $P(x)$ is unknown and hence this not simply a matter of finding equation roots. The procedure is sequential and will converge to x_{100p} under the conditions for $P(x)$ given by (1.1). For this reason it is also considered an estimation technique, as indicated in Section 1.4.1. Many variations of the Robbins-Monro (RM) procedure exist. See, for example, Kesten [1958], Anbar [1978] and Lai and Robbins [1979]. However, we will discuss only a version of the RM strategy for which convergence is delayed. This particular strategy has performed well in Monte Carlo studies involving small samples. See, for example, Cochran and Davis [1964], Davis [1971], and Bodt and Tingey [1986].

The Delayed Robbins-Monro (DRM) procedure of Cochran and Davis [1964] selects design points converging to x_{100p} as follows. Denote the i^{th} level of stimulus as x_i with observation y_i, where $y_i = 1$ signifies a response and $y_i = 0$ signifies a nonresponse. The next design point x_{i+1} *for a DRM design is given by*

$$x_{i+1} = x_i - c\,(y_i - p),$$

where c is an appropriately chosen constant according to the variance of the population. Data is collected in this manner until a reversal of response type is observed in successive trials. Subsequent design points are chosen according to a usual form of the Stochastic Approximation Method as

$$x_{i+1} = x_i - \frac{c}{i - k + 1}\,(y_i - p),$$

where k is the first sample corresponding to the first reversal. The delay causes the design to refrain from attempted convergence until some indication (reversal) of being in an appropriate range of the stimulus is present. If starting in the tail of the response distribution, immediate attempted convergence would be unwise, particularly with small sample sizes.

1.4.3 Some Properties of Sequential Estimation Procedures

Though our interest is extreme quantile estimation, it is important to note the performance of median estimators as they compare to estimators for general x_{100p}. In this section emphasis is given to properties of median estimators. The relevance of those properties to extreme quantile estimation is discussed in Section 2.3.

The properties surrounding x_{50} estimators are well known through theoretical and Monte Carlo investigations. Of particular interest, are those estimation procedures involving sequential design. The results indicate that most of the common estimation procedures yield estimates of x_{50} which are accurate, precise, and robust under the usual parametric assumptions.

Accurate estimation of x_{50} is possible with large or small samples. For large samples, consistent estimation of the more general x_{100p} is achieved by either the RM procedure or by maximum likelihood estimation provided the parametric form assumed is correct. For small samples, Monte Carlo studies have shown estimate unbiasedness for symmetric distributions. See, for example, Wetherill [1963] and Davis [1971]. Bodt and Tingey [1986] demonstrated that good small sample estimation is still possible when the distribution is asymmetric. Specifically, for the exponential distribution the mean square error associated with $\hat{x}_{50}$ was comparable to that of three symmetric response distributions. Their estimation procedure consists of collecting data according to the DRM procedure and estimating using maximum likelihood with an assumed normal distribution.

In contrast, although asymptotically unbiased estimates for extreme x_{100p} are possible, small sample estimates are generally biased. Wetherill [1963] argued that biased extreme quantile estimates resulted from small sample application of the RM procedure. Wu [1985] and Bodt and Tingey [1987] showed small sample bias for more recent estimation procedures, including those given by Anbar [1978] and Wu [1985].

The precision associated with x_{50} estimates is also better than the precision associated with extreme x_{100p} estimates. For example, we consider precision with respect to the Robbins-Monro strategy. There the estimate of x_{100p} has an asymptotic variance proportional to $1/\{p(1-p)\}$, the variance achieving its minimum at $p = .5$. In small sample application, the observed precision agrees well with the asymptotic results; Wetherill [1963] demonstrated empirically an approximate 80% efficiency for a sample size of fifty. However, for extreme quantiles, variances much larger than their asymptotic values resulted from the small sample application of this procedure [Wetherill, 1976]. Numerous other simulation studies support the "better precision" claim over a variety of parametric forms, experimental conditions, and data collection procedures. See, for example, Rothman, et al. [1965], Hsi [1969], Wu [1985], and Bodt and Tingey [1987].

Robustness to parametric form is a well established property of x_{50} estimators. "If testing and conclusions are confined to a region near the 50% point, x_5, then the experimenter can hardly go wrong with any model he uses" [Rothman et al. 1965]. Davis [1971] demonstrated robustness in a Monte Carlo study for several procedures including the DRM. Little [1974] also showed this property for some mildly skewed parametric forms. Thus we may claim some freedom in the selection of a parametric form for the estimation of x_{50}.

2. SEQUENTIAL PROCEDURES FOR EXTREME QUANTILES

Sequential methods serve as the basis for many extreme quantile (extreme value) design and estimation procedures. The popularity of these sequential approaches stems from their attractive tendency to converge or restrict sampling to the region of interest, that is, near x_{100p}. This aspect need not be linked to restrictive assumptions, which is especially important when estimating in the tail of a response distribution. There, parametric assumptions are often conjecture; consequently, even the general location of x_{100p} may be unknown. Additionally, their economy-of-subjects property facilitates experimentation when subjects are expensive. In Section 2.1 we discuss some prevailing sequential procedures intended specifically to estimate extreme quantiles, and in Section 2.2 some issues regarding their use. Lastly, in Section 2.3 we introduce a new procedure and argue its attractive characteristics with respect to the issues raised in Section 2.2.

2.1 Current Methods

The sequential aspect of these methods refers to a one-subject-at-a-time application of stimulus level. If only one subject is tested before selection of a new stimulus level we call the procedure sequential. The term block sequential indicates that additional subjects may be tested before moving on. In Sections 2.1.1-2.1.3 we discuss sequential and block sequential procedures. In Section 2.1.3 we consider the notion of a transformed response. Strictly speaking, methods based on transformed responses belong to the block sequential class of designs; they appear separately and in more detail because of their importance to our approach.

A discussion of estimation using sequentially collected data appears in Section 2.1.4.

2.1.1 Sequential Design

Straightforward binomial-based arguments support the use of most extreme value sequential designs. For instance if we seek an estimate of x_{95} then a stimulus level at which nineteen of twenty subjects respond is of obvious interest. A sequential design which uncovers such a stimulus level not only has provided the basis for a reasonable nonparametric point estimate but probably also has, in the process, collected data about the true x_{95}. Using such data for parametric estimation results in the avoidance of a major pitfall of extreme value estimation--extrapolated estimates. To address the task of collecting data in a region containing x_{100p} many sequential schemes play off the binomial theme.

McLeish and Tosh [1983] offered a representative design with simple rules. As with the Up and Down method of Section 1.4.2, they considered equally-spaced levels of stimulus as potential design points. To collect data relevant to the estimation of x_{05} their procedure calls for an initial design point selection thought well below the x_{05}, where a nonresponse is the likely result. The design chooses the next highest stimulus for subsequent design points until the sequence ends with the observance of the first response. The process may be repeated, yielding many of these sequences which individually and collectively hold information about the lower tail of the response distribution.

For individual sequences, confidence that the design will collect meaningful data is gleaned from a simple binomial exercise. Suppose that the true response function is normal(μ, σ), the initial design point is $\mu - 3\sigma$, and the spacing between levels is $.5\sigma$. Recognize that each observation in a sequence results from an independent Bernoulli trial. With the probability of a nonresponse known for each design point, we easily compute the probability to be .74 that a sequence does not wander beyond the median. Thus the preponderance of information gathered from repeated sequences concerns a gross region of interest--the lower half of the response distribution. Judicious stimulus spacing and initial design point selection allow for design focus on a more specific region. Thus we can reasonably ensure each sequence will gather useful information.

The sequences possess a collective utility through shared stimulus levels. A stimulus level shared by n sequences supports n identically distributed and independent Bernoulli trials. These replicate observations certainly benefit estimation, and in doing so they support more involved sequential designs such as the Alexander Extreme Value Design [Rothman et al. 1965]. There, binomial probabilities associated with n outcomes at some level were used to establish stopping rules addressing the number of sequences needed.

Designs discussed by Rothman et al. [1965] include the Naval Powder Factory (NPF), the Alexander Extreme Value, and the Rothman. The NPF and Alexander Extreme Value designs are similar to the design examined by McLeish and Tosh [1983] in that a sequence is formed by choosing an adjacent level, among a group of fixed equally-spaced levels, to be the next design point. They differ from this design in that they both employ alternating increasing and decreasing sequences with stopping rules, though different, both in keeping with the binomial-based arguments mentioned above.

For example, consider application of Alexander's procedure to finding x_{05}. Denote the spacing constant by δ and the lowest level at which a response is observed by x^R.

The alternating increasing and decreasing sequences end when n nonresponses and 0 responses have been collectively observed on $x^R - \delta$ and $x^R - 2\delta$. The most important feature of this design is this stopping rule. If sampling at the true x_{05}, the probability p that at least one response occurs out of n trials is given by $1 - (1-.05)^n$. Since the response curve is increasing, p represents the minimum probability of observing at least one response out of n trials for any fixed level of stimulus on $[x_{05}, x_{100}]$. Failure to observe any responses suggests, with confidence afforded by the magnitude of p, that the stimulus being tested is actually below x_{05}. To select n we merely choose a value for p and solve. In practice, the authors suggested considering the nonresponses at $x^R - \delta$ and $x^R - 2\delta$ as all coming from the latter stimulus. Then the conclusion, with n nonresponses at those two levels, is that this latter stimulus is below x_{05}.

An exception to the binomial-based arguments is found in the justification of the Rothman design. The Rothman design uses no fixed levels or alternating increasing and decreasing sequences. Rather, the next design point is derived from a constrained maximum likelihood procedure discussed by Ayer et al. [1955]. The constrained maximum likelihood procedure is more appropriately discussed in Section 2.1.4. Of note is the idea of using MLEs based on the first n data points in the selection of the next design point. This general idea is again implemented in a sequential procedure suggested by Wu [1985] for estimation of the median tolerance.

2.1.2 Block Sequential Design

Block sequential designs provide another means for data collection. Their justification coincides with that of sequential designs in as much as binomial-based arguments support the use of both. They differ from sequential designs in that replicate observations receive greater emphasis. This emphasis is achieved by requiring, in most cases, replicate information to be incorporated in the procedure's selection of the next design point. Replicates at the current stimulus level convey, through the sample response probability, greater information regarding the design's present position relative to x_{100p}.

Bartlett [1946] offered a design representative of several in this class. As with other procedures the design moves among a set of equally-spaced test levels. To gather information for small values of p, testing begins at a stimulus level thought to be near x_{50}. Testing continues there until two responses have been observed, at which point the design drops down to the next lower level of stimulus. At this next lower level the same two-response rule applies, and so on. Though more economical in terms of subjects than most fixed designs, the Bartlett procedure requires many samples if information regarding the extreme tail is desired. For instance, if testing at x_{05} we expect forty samples to be required for two responses at this stimulus level alone.

The n-Zill design, discussed by Rothman et al. [1965], protects against the collection of an excessive amount of data at one level with a stopping rule; if n nonresponses are observed before the first response, the procedure ends. The choice of n determines how far out in the tail 'he design is likely to move. This design begins as the Bartlett design does, but seeks only the first response before moving down. Additionally, it skips a level if the response occurs in the first five trials. An analogous design for either is possible for the upper tail of the response distribution.

2.1.3 Transformed Response Design

Procedures using the transformed response rule of Wetherill [1963], studied in detail by Wetherill et al. [1966], belong to the block sequential class; however, they differ somewhat from other members of this class in their interpretation of outcomes. Consider a conceptual fixed sample size n_c for a given block of observations at a given stimulus. A basic transformed response strategy partitions the 2^{n_c} possible outcomes into two sets, denoting one set a success and the other a failure. Success and failure represent response and nonresponse, respectively, in the context of the transformed response distribution. (Hereafter, success and failure always refer to transformed responses, and response and nonresponse always refer to original outcomes.) Success or failure alone, not the original outcomes themselves, determines the direction, up or down, for continued sampling. The advantage of this approach is that it allows the experimenter to collect data and estimate in terms of the transformed response distribution.

First we illustrate the mechanics of a transformed response strategy for data collection, with the estimation argument to follow. Consider the conceptual sample size n_c to be three; and define a success as the set { 1 1 1 }, where 1 signifies a response and 0 signifies a nonresponse. Choose as the sequential strategy the Up and Down method described in Section 1.4.2. Here the design moves down one fixed equally-spaced level upon observance of a success and up one level for each failure. Figure 2.1 summarizes representative results for ten blocks, with stimulus level serving as the ordinate. The response/nonresponse ordering corresponds to the testing order of subjects within a block. The third subject in the first block failed to respond. Consequently, the block result { 1 1 0 } is classified a failure. The next block, taken at the next highest level, yields a success { 1 1 1 }. Note that for some blocks we determine a failure with less than three observations. This is why we refer to n_c only as a conceptual sample size.

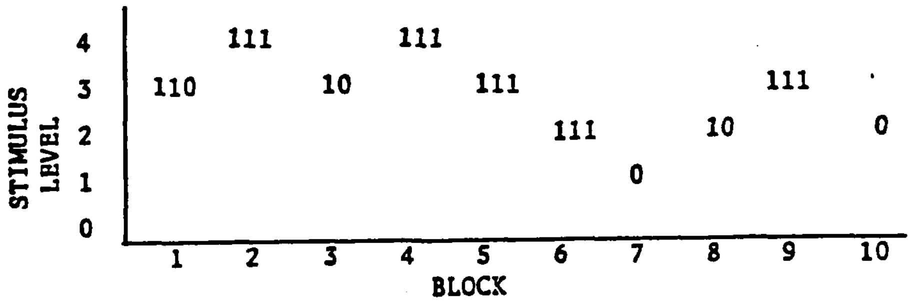

Figure 2.1 Representative results for ten blocks

To justify estimating quantiles of P(x) by estimating quantiles of the transformed response distribution $T\{P(x)\}$ we need to

1. determine conditions for $T\{\cdot\}$ such that $T\{P(x)\}$ is a distribution function,

2. show that the quantiles of P(x) can be expressed in terms of the quantiles of $T\{P(x)\}$, and

3. provide motivation for estimation in terms of $T\{P(x)\}$.

First, we require $T\{\cdot\}$ to be a continuous monotone increasing function on [0,1], with $T\{0\} = 0$ and $T\{1\} = 1$ to ensure that $T\{P(x)\}$ is a distribution function. The strictly-increasing condition may be viewed as a modeling convenience. Second, a monotone increasing function is one-to-one, thereby guaranteeing the existence of an inverse. Then through its inverse, complete knowledge of $T\{P(x)\}$ constitutes complete knowledge of P(x). Third, a judicious selection of $T\{\cdot\}$ may allow for the estimation of quantiles more tractable in terms of $T\{P(x)\}$ than in terms of P(x).

The following illustrates the role played by the above three issues. Return to the application of the transformed response strategy in this section. There $T\{\cdot\}$ took the form $T\{P(x)\} = P(x)^3$, where $P(x)^3$ is the probability of success for a stimulus level x, or equivalently the probability that each of three subjects registers a response at this level. Clearly $P(x)^3$ satisfies the conditions set forth above. As a sidelight, note that $P(x)^3$ is the distribution of the maximum order statistic from a random sample of three tolerances. Now let t_{100q} denote the q^{th} quantile of $T\{\cdot\}$, and let x_{100p} be the stimulus level for which $P(x) = t_{100q}$. Through the inverse relationship $T^{-1}[T\{P(x)\}] = P(x)$ we have

$$T^{-1}[T\{t_{100q}\}] = q^{1/3} = P(x_{100p}).$$

Thus, the stimulus x_{100p} giving rise to the q^{th} quantile of the transformed response curve is the $q^{1/3}$ = pth quantile of P(x). For this example note that the median of $T\{P(x)\}$ corresponds to the $.5^{1/3} = .7937$ quantile of P(x). The value .7937 is termed the transformed median quantile. Thus we can acquire information regarding the extreme tail of P(x) through the accurate, precise, and robust estimation of the median, discussed in Section 1.4.3, of the transformed response distribution.

The basic transformed response strategy described above consists of defining a success in such a way as to transform P(x) by raising it to a positive integer power. We may define a success in other ways leading to different transformations so long as the conditions for $T\{\cdot\}$ are satisfied. In fact, a success need not be a partition of just

2^{n_c} possible outcomes. Consider a strategy in which $n_c = 3$, { 1 1 1 } is a success, { 1 1 0} is indeterminate as to success or failure, and all other outcomes are classified as failures. In the case of the indeterminate outcome, the strategy requires an additional sample; a response yields { 1 1 0 1 } which is classified a success and a nonresponse leads to the failure { 1 1 0 0 }. Thus the strategy dictates a probability of success and the transformation $T\{\cdot\}$ given by

$$T\{P(x)\} = P(x)^3 + P(x)^3(1 - P(x)) = P(x)^3(2 - P(x)).$$

Some strategies, their corresponding transformations, and their transformed median quantiles are given in Table 2.1.

Recognize two facts pertaining to Table 2.1. First, only a limited number of P(x) quantiles appear with the strategies given. We address this point further in Section 2.2. Second, we may derive strategies for the lower tail of the response distribution by

1. reversing the roles of 0 and 1 to denote response and nonresponse, respectively, and

2. reversing the actions associated with success and failure.

For example, let us employ the Up and Down strategy for small p with $n_c = 2$. Upon observance of { nonresponse, nonresponse }, a success, the design moves up one level--gathering information for the transformed median response .2929, that is, the value of p which is the solution of $(1 - p)^2 = .5$.

Although many designs are appropriate for use with transformed responses, only two appear in the literature. The Up and Down method acting on transformed responses (UDTR) was introduced by Wetherill [1963]. Einbinder [1973] suggested implementing Langlie's [1962] One Shot strategy on transformed responses (OSTR). The One Shot strategy may be thought of as a variable step size Up and Down approach. Robbins-Monro based designs have not been used with transformed responses. This is unfortunate because RM based designs have been found to be superior performers in numerous Monte Carlo investigations. Use of an alternative design in conjunction with transformed responses constitutes a portion of the method introduced in Section 2.3.

2.1.4 Estimation for Sequential Procedures

In this section we discuss the estimation techniques proposed for use with the above sequential procedures. Usually, more than one estimation technique fits well with each design, but all of the designs draw from among the same few estimation

Table 2.1. Transformed Response Strategies

n_c	Success	Failure	Transformation	Transformed Median
2	11	10, 0	p^2	.7071
3	111, 1101	1100, 10, 0	$p^3(2-p)$	.7336
3	111	110, 10, 0	p^3	.7937
4	1111, 11101	11100, 110, 10, 0	$p^4(2-p)$	.8041
4	1111	1110, 110, 10, 0	p^4	.8409
5	11111, 111101	111100, 1110, 110	$p^5(2-p)$	.8460
5	11111	11110, 1110, 110, 10, 0	p^5	.8706
6	111111	111110, etc.	p^6	.8909
7	1111111	1111110, etc.	p^7	.9057
8	11111111	11111110, etc.	p^8	.9170
9	111111111	111111110, etc.	p^9	.9259
10	1111111111	1111111110, etc.	p^{10}	.9330
14	11111111111111	11111111111110, etc.	p^{14}	.9517

NOTE: In this table p(x) is denoted p.

techniques. For this reason we structure the section according to the principal estimations used, noting for which designs the estimations are appropriate. For discussion purposes interest rests in the lower tail of the response distribution.

The next lowest stimulus $x^R - \delta$ below the lowest stimulus yielding a response x^R gives a rough nonparametric estimate for lower threshold values. This estimate depends entirely on the rationale underlying the design implemented. Specifically, it relies on both stopping rules to determine what will be considered the lowest response stimulus and the step size δ. For example, consider the n-Zill design. Recall that the decreasing sequence stops at the first level in which n nonresponses have been observed without a response. Let us estimate the lower threshold with $x^R - \delta$. Sampling at x_{100p}, the probability that the strategy chooses x_{100p} to serve as $x^R - \delta$ is given by $(1 - p)^n$. Thus the choice of n greatly influences the estimate value $x^R - \delta$. The selection of n, that is, choice among practicable n-Zill designs stochastically determines the region in which $x^R - \delta$ is likely to fall. The spacing δ has fairly obvious consequences. A δ too small results in a more refined estimate but at the expense of additional samples likely to be necessary for a greater number of levels considered. A δ too large results in a less refined estimate. Informed selection of δ requires knowledge, usually unknown, of the scale of the response distribution.

Designs appropriate for use with $x^R - \delta$ include all of the sequential procedures of this chapter, maybe with minor changes, except possibly the Rothman and OSTR procedures. The requirement for use is only that $x^R - \delta$ is a reasonable estimate in consideration of the design behavior. By reasonable we mean that a stopping rule for a sequence or a number of sequences is likely, according to the Bernoulli response probabilities, to result in a $x^R - \delta$ close to the quantile of interest. Minor changes entail the creation of such stopping rules for designs which have none except for sample size limitations. The reason for excluding the Rothman and OSTR is that their variable step sizes dictate an uncertain distance below x^R for the estimate value. For each of the other designs stimulus levels are equally spaced.

The estimation procedure discussed by Ayer et al. [1955] yields MLEs for the response probability associated with each stimulus level tested. Denote this probability $P(X_i)$, and assume that

$$P(x_1) \geq P(x_2) \geq \cdots \geq P(x_n) \tag{2.1}$$

for decreasing levels x_i, $i = 1, 2, \cdots, n$.

For stimulus x_i denote the number of responses by r_i and the proportion of responses by $P(x_i)^* = r_i/n_i$. The MLEs $P^\nabla(x_1), P^\nabla(x_2), \cdots, P^\nabla(x_n)$ for $P(x_1), P(x_2), \cdots, P(x_n)$ are assigned as follows. If the sample proportions $P(x_i)^*$ conform to the constraint expressed in (2.1), then set $P^\nabla(x_i) = P(x_i)^*, i = 1, 2, \cdots, n$. If $P(x_i)^* \leq P(x_{i+1})^*$ for some $i = 1, 2, \cdots, n-1$, then set $P^\nabla(x_i)$ equal to $P^\nabla(x_{i+1})$, and compute their common value as the ratio $(r_i + r_{i+1})/(n_i + n_{i+1})$. This single new sample ratio replaces $P(x_i)^*$ and $P(x_{i+1})^*$ in the sequence leaving n - 1 ratios. If the sequence of n - 1 remaining ratios conforms to the initial constraint, we may stop. If not, repeat the procedure until the desired ordering is obtained.

The above algorithm ensures finding MLEs for certain response probabilities, but it does not directly address estimation of specific quantiles chosen in advance. Rothman et al. [1965] employed linear interpolation for this task. This general procedure is recommended for use with the Rothman design and the Alexander Extreme Value design, but it could be used with other designs of this chapter. We illustrate the technique with partial results from an Alexander Extreme Value design.

Suppose we wish to estimate x_{05} using the following data. The stimulus levels $x_i, i = 1, 2, \cdots, 5$ correspond to $\{2, 1, 0, -1, -2\}$, and the ordered set $\{\frac{1}{2}, \frac{0}{2}, \frac{1}{6}, \frac{0}{7}, \frac{0}{9}\}$ are the respective $P(x_i)^*$'s. Since $P(1)^* < P(0)^*$ we let $P^\nabla(1) = P^\nabla(0) = 1/8$ to satisfy the order constraint and arrive at MLEs for the response probabilities. Linear interpolation between $P^\nabla(0)$ and $P^\nabla(-1)$ with respective probability estimates 1/8 and 0 yields an estimate stimulus level of -.6 for the x_{05}.

Maximum likelihood estimation is the most commonly used method for estimating extreme quantiles. If we assume a two-parameter family of distributions we may proceed as in Section 1.4.2, solving (1.2) for α and η and (1.3) for the quantile of interest. Extreme quantiles were historically estimated in this fashion, usually with a normal or logistic distribution assumed for $P(x)$. The designs of this chapter all produce data suitable for use with this technique.

Two computational considerations exist with this approach. First, the estimates need not exist for each set of data. Conditions guaranteeing MLE existence are discussed in detail in Chapter 3. Second, generally (1.2) cannot be solved directly for the parameter estimates as is the case for normal and logistic assumptions. Thus, we must rely on iterative schemes such as the Newton-Raphson process or the Method of Scores. Either method, if multiple roots exist, may converge to a root which does not correspond to the maximum. Alternatively, either may fail to converge to any root. To the latter issue DiDonato and Jarnagin [1972] offered an iterative approach guaranteeing convergence to the global maximum under the normal parametric assumption. Data should be collected with both of these considerations in mind.

Parametric forms with greater than two parameters may also be used. Einbinder [1973] used a three parameter Weibull distribution for $P(x)$. He collected data according to the OSTR strategy and then formed MLEs of the three parameters. Finding estimates in the three parameter case can be more complex when iterative procedures are necessary. One method involves searching over a reasonable parameter space for one estimate, optimizing at each point with respect to the other two parameters. Justification for parametric forms such as the Weibull was mentioned in Section 1.4 and emphasized in Section 2.2.

McLeish and Tosh [1983] estimated in terms of a first-response distribution. Recall that they proposed data collection in increasing sequences of equally-spaced stimulus levels until the first response. The first-response stimulus x_N depends on the initial stimulus x_1, the stimulus spacing δ, and the range of stimulus levels $(N - 1)\delta$. Assume that the response distribution P(x) is logistic with parameters τ and ω;

$$P(x) = (1 + e^{-\omega(x - \tau)})^{-1}.$$

They could, by maximum likelihood, have estimated τ and ω directly by processing each stimulus/response data point through (1.2). Instead they chose to first summarize the information from each sequence in terms of x_1, δ, and $(N - 1)\delta$, and then to estimate shared parameters according to the first-response distribution.

Given x_1 we need only the range of doses $(N - 1)\delta$ to determine the first response. After making a continuity correction of $\delta/2$ they approximated the distribution of $D = (N - 1/2)\delta$ with a continuous distribution. Specifically, for $\delta \rightarrow 0$ and $x_1 \rightarrow -\infty$, $e^{\omega D} - 1$ has an approximate exponential distribution with mean $1/\lambda$, where

$$\lambda = \frac{e^{\omega(x_1 - \tau)}}{e^{\omega\delta} - 1}.$$

The approximation is good when $x_1 \leq x_{10p}$ and $\delta \leq .1\omega$ [McLeish and Tosh 1983]. From the experiment, x_1 and δ are known, and the parameters τ and ω, shared between the logistic and first-response distributions, may be estimated. Realizations of D are processed through the log-likelihood equations from the approximate first-response distribution to yield MLEs for τ and ω. Then for small p, $\hat{x}_{100p}$ for the assumed logistic response distribution is computed as a function of $\hat{\tau}$ and $\hat{\omega}$. It is important to note that this estimate is not computed directly from the data, but through a summary of the data.

Wetherill et al. [1966] explored the use of $\overline{w}$ as an estimator for the UDTR strategy. In the usual implementation of the Up and Down strategy, there exist several pairs of successive levels for which the response changes, that is, reversals occur. Define w_i as the average stimulus level corresponding to the i^{th} pair. Only these values w_i holding reversal information are averaged together to form $\overline{w}$, an estimate of the median. For transformed responses, reversals depend on success and failure. Referring to Figure 2.1 in Section 2.1.3, the w_i occur as $\{3.5, 3.5, 3.5, 1.5, 2.5, 2.5\}$, and they are averaged to form $\overline{w} = 2.83$. Its interpretation in terms of P(x) is given in Table 2.1 as the transformed median quantile .7937.

The estimator $\overline{w}$ is based on response type as well as the stimulus levels tested. When Dixon and Mood [1948] introduced the Up and Down method, they included a simple estimator. With a normality assumption they showed that the response distribution parameters may be estimated using maximum likelihood. However, the solution must be arrived at iteratively. To overcome this computational inconvenience they proposed a simple technique of stimulus level averaging to approximate the location estimate. Brownlee et al. [1953] followed with alternative dose-averaging methods which took into account when in the sequence each stimulus was tested; for instance, in one average they excluded the first stimulus level tested, claiming that since it was chosen by the experimenter it did not contain information about the location parameter. The $\overline{w}$ estimator screens the data further through the consideration of response type. A reversal of response type indicates with limited certainty that the two stimulus levels involved straddle the response distribution median. Thus each average w_i can be viewed as an estimate of this median with $\overline{w}$ serving as a composite estimate. Besides the intuitive appeal, support for $\overline{w}$ relies primarily on its favorable Monte Carlo performance relative to other dose-averaging techniques [Wetherill et al. 1966].

Note that using $\overline{w}$ with the UDTR strategy constitutes a departure from the conventional estimation procedures for extreme quantiles. The estimated median belongs to the transformed response distribution and not to the original tolerance distribution. Consideration of an alternative estimator under transformation comprises a portion of the new proposal in Section 2.3.

2.2 Critique of the Issues in Estimation

In the previous sections, while introducing various techniques, we have touched on several of the issues in extreme quantile estimation. In this section these issues are discussed in more detail. We structure the discussion around the type of estimation procedure employed, noting specific design considerations where appropriate. The estimation procedures are considered to be of three types: nonparametric, parametric maximum likelihood, and summary information. Summary information refers to transformed response strategies and the first-response approach. For each estimation procedure we discuss problems and literature attempts to address those problems. The discussion of this section is intended to lay groundwork for the new procedure given in Section 2.3.

2.2.1 Nonparametric Procedures

In this section nonparametric procedures include the next lowest stimulus below the lowest stimulus yielding a response, $x^R - \delta$, and the constrained maximum likelihood estimation of Ayer et al. [1955]. The interest in a nonparametric approach stems from the limited knowledge experimenters have regarding the tail of the response distribution. Remember that one reason sequential procedures are desirable is that even the general location of x_{100p} is unknown. Nor is the parametric form known, at least for the distribution tail. Nonparametric estimation overcomes this problem, though possibly at some expense. The two nonparametric techniques share three primary concerns, namely, stimulus level spacing, sample size requirements, and the practicability of inference beyond simply the point estimate for the quantile of interest.

Large gaps between adjacent stimulus levels lead to an estimate which may only roughly approximate the quantile of interest. Consider the sequential designs where δ is the equal spacing between stimulus levels. A large value of δ relative to the response distribution standard deviation may

1. inflate the root mean squared error associated with $x^R - \delta$ depending on the actual stimulus level placements, and

2. prevent collection of meaningful data for *constrained maximum* likelihood estimation.

To see both let the response distribution be normal with mean μ and standard deviation σ. Let the potential stimulus levels be taken from $\mu - \sigma \pm 2k\sigma$, $k = 0, 1, 2, \cdots$, and estimate $x_{.023}$ corresponding to the stimulus level $\mu - 2\sigma$. Note that since the potential stimulus levels exactly straddle $x_{.023}$, the nearest stimulus level possible is $\sigma = \delta/2$ distance away. Then the root mean square error associated with $x^R - \delta$ can be no smaller than $\delta/2$. Alternatively, suppose that the constrained maximum likelihood approach dictates a linear interpolation between $\mu - \sigma$ and $\mu - 3\sigma$. Linear interpolation of the asymptotic values of the estimates yields an estimate of $\mu - 1.4\sigma$. Arguments follow similarly for stimulus levels with variable spacing.

One can argue that the above example is contrived and δ and stimulus level placements need not be so poorly chosen. However, rational choices for each require information about the unknown scale parameter and the unknown quantile of interest. Many authors point this out including Wetherill [1963], Rothman et al. [1965], and Hsi [1969]. They each presented Monte Carlo evidence to suggest appropriate choices for each, but the choices were expressed as a function of the two unknowns. Preliminary sampling and parametric estimation could provide initial values for location and scale.

However, in keeping with the nonparametric intent, achieving better estimates requires improving the stimulus spacing irrespective of parametric estimates of the unknowns.

To address this task the spacing between levels may be gradually decreased based on the sequentially gathered information, or the spacing may be intentionally chosen to be narrow so that a finer resolution of information is obtained. Potentially this creates another problem, increased sample sizes. Consider that narrow spacing is likely to increase sample sizes over the entire range of stimulus levels tested, but the nonparametric estimates of this section draw upon only one or two of these levels to make their estimates.

Sample size is an important consideration when using these nonparametric estimation procedures. In most experimental environments we are limited in some way with respect to sample size. The limitation is usually expressed in terms of a cost such as time or number of units destroyed. The latter was partial motivation for the first-response approach. This cost is offset by some measure of information gain. Above we note that narrow spacing is likely to result in an increased number of samples. Additionally, large sample sizes are required to satisfy the stopping rules which we mentioned. We say the sample sizes needed are large because the stopping rules depend on the estimation of very small or very large probabilities with binary data. Thus, if we wish to adequately estimate some target quantile using a nonparametric approach, we must be prepared to collect a large number of samples.

Another concern is that the nonparametric estimators only provide reasonable point estimates for one specific quantile of interest. Although in many applications this may be sufficient, more complete information regarding the response distribution would be useful. For example, a chemical test is designed to respond to measurable quantities of some substance. The response probability increases with the quantity of substance. For the purpose of establishing a reliability standard the experimenters need to know the quantity of substance corresponding to a response probability of .85. A nonparametric approach will allow for such a determination, but in this case it is also reasonable to explore the behavior of the chemical test in a region about x_{85}. Perhaps the ordered estimates in the constrained maximum likelihood approach can lend some insight in this regard, but there only some bounding can be accomplished. A procedure capable of estimating x_{85} and neighboring quantiles is more desirable.

The performance of these estimators is open to question. Rothman et al. [1965] claimed that nonparametric estimation for the Alexander Extreme Value design performs about as well (having approximately equivalent root mean squared error) as parametric procedures using the true parametric assumption. The parametric procedures they referred to consist mainly of invoking maximum likelihood estimation on data collected by a variety of sequential designs including the Bartlett and n-Zill.

However, the Monte-Carlo study supporting this claim was limited to one hundred iterations, and the exact manner of performance comparison with regard to equal sample sizes, etc., is not clearly stated.

2.2.2 Parametric MLE Procedures

The issues concerning parametric maximum likelihood estimation of extreme quantiles include response distribution assumptions, design, limited resources, and computational factors. Much of the extreme value literature consists of attempts to resolve problems involving these factors, although not exclusively in consideration of maximum likelihood estimation. In this section, we discuss these issues in a chronology roughly paralleling their treatment in the literature. The time-ordered presentation also provides a convenient framework for relating, according to these issues, the designs of Section 2.1.

Prior to the interest in extreme quantile estimation, the favored response distribution assumptions were the normal and logistic distributions. Bliss [1934a,b] introduced the normal response assumption for use in bioassay. Some practical applications in which the response curve has been extensively studied support this assumption. However, more generally "the central limit theorem gives reason for hoping that conclusions based on the normal assumption will be close to the truth when means of several observations are involved" [Finney, 1978]. Berkson [1944] argued on behalf of the logistic assumption, citing its similarity to the normal assumption and its greater mathematical tractability. Other forms considered include the uniform, Cauchy, and angle ($\sin^{-1}\sqrt{p}$) transformations. Since they all closely approximate the normal over (.25, .75), shape was not a serious discerning factor in choosing among them. Borne out in later studies, cited in Chapter 1, many median estimators, including maximum likelihood, prove robust among these and other response distribution assumptions. This fact deprives the distribution assumption issue, in terms of median estimation, of any practical significance.

The other three issues--design, resources, and computational factors--stimulated little debate. The probit approach, Bliss [1934a,b], became the standard in design and analysis for studying dose response curves. Since much of the early work focused on biological applications, data was readily available. The common design with probit analysis involved sampling many stimulus levels over the practicable stimulus range, gathering multiple observations at each level. Analysis consisted of an iterative formation of linear regressions meant to bound the true linear relationship assumed to exist between the stimulus and $F^{-1}(p)$. Here $F(\cdot)$ represents a normal distribution with mean 5 and variance 1. The mean value was selected to avoid potential confusion possible with negative values of the stimulus; the stimulus is often a necessarily positive term such as drug dosage. Usually, a couple of iterations sufficed; and they could be accomplished graphically, providing no great computational concern.

The issues of this section became more important in the late 1940s concurrent with a wider application of sensitivity analysis. In the physical sciences, the limited resources and interest in extreme quantiles motivated Bartlett [1946] to suggest the design given in Section 2.1.2. Bartlett recognized that it was important to sample near the quantile being estimated, and he suggested doing so with a sequential strategy. Robbins and Monro [1951] further advanced the concept of sampling about a general desired quantile with their Stochastic Approximation Method. All of the strategies given in Sections 2.1.1-2.1.3 adhere to this basic idea, citing binomial probability arguments as support.

Justification in terms of maximum likelihood estimation for sampling in the region of interest was not given formally until 1962. Chernoff [1962] examined, for the normal assumption, the asymptotic variance of $\hat{x}_{100p}$. He developed optimum fixed designs which minimize $\mathrm{Var}(\hat{x}_{100p})$, or $\mathrm{Var}(\hat{\mu} + Z_p\hat{\sigma})$, where the inverse information matrix provides the necessary variance and covariance values for $\hat{\mu}$ and $\hat{\sigma}$. For quantiles x_{100p} in the range x_{06} to x_{94} the strategy suggests allocating all samples to the stimulus corresponding to x_{100p}. For quantiles outside this range the design selects two stimulus levels $\mu - 1.57\sigma$ and $\mu + 1.57\sigma$ in proportions $Z_p - 1.57$ to $Z_p + 1.57$. His results suggest, except in the case of quantiles outside (x_{06}, x_{94}), that when using maximum likelihood estimation with a normal assumption, data should be collected in the neighborhood of the target quantile.

The practical application of this design is difficult for two reasons. First, the optimum stimulus level selections depend on the unknown parameters μ and σ. Chernoff [1962] suggested using a preliminary design to estimate μ and σ. These estimates are substituted for the true parameters in the design point selection. Second, for quantiles between x_{06} and x_{94} the design samples at only one level of stimulus. This results in a failure to meet the existence conditions for the MLEs, thus preventing estimation of x_{100p}. A partial design solution to these problems exists in the sequential strategies of Sections 2.1.1 and 2.1.2. We discuss in Chapter 3 how one of those strategies can be used to overcome these problems.

Much of the preceding discussion focuses on design and estimation when a normal response function can be assumed. However, rarely if ever are distribution behaviors known to the extent that distributional assumptions can be made in consideration of the tail. Thus difficulties in estimation with an incorrect parametric assumption may result. Two methods of accounting for the uncertain parametric form are

1. estimate without benefit of a specific parametric form, or

2. estimate with a parametric form considered robust among many possible parametric families.

Many of the procedures discussed in Section 2.1, not requiring parametric maximum likelihood, approach the problem according to method one. We discussed issues concerning the use of early nonparametric techniques in Section 2.2.1, deferring discussion of the UDTR strategy issues until Section 2.2.3. Several researchers have proposed robust parametric forms or stimulus transformations in the sense of method 2. See Einbinder [1973], Prentice [1976], Little [1976], Copenhaver and Mielke [1977], Egger [1979], Aranda-Ordaz [1981], Guerrero and Johnson [1982] and Morgan [1985]. All models possess three or four parameters and include forms similar to the logistic model as a special case.

The argument for preferring one of these robust parametric forms follows. First, these robust families are able to emulate the common logistic model. Their performance relative to maximum likelihood under logistic and normal assumptions appears to be good. Second, except for the distribution of Copenhaver and Mielke [1977] the response distributions may assume asymmetric forms. Third, the shape of the tail of the distribution is more flexible--thought to be important when the quantile to be estimated is an extrapolation of the data. Thus, the new distributions constitute a more general class of the distributions already used.

A selection from among these robust parametric forms must take into account the following points. The design influence on estimation with these distributions has not been studied. Instead, their performance on available data sets serves as a basis for comparison. Also, none of the distributions proposed have distinguished themselves relative to the others. In light of this, the practical concern of computational ease becomes an issue. Estimation in each case requires good computational facilities. Some require Newton-Raphson iterative solutions for a three parameter model while others require the use of a numerical algorithm, GLIM [Baker and Nelder 1978].

2.2.3 Summary Information Procedures

Summary information was alluded to in Section 2.1 as data condensed from the raw quantal form. Approaches to extreme quantile estimation given by Wetherill [1963] and McLeish and Tosh [1983] each use summary information. In this section we discuss issues concerning design and estimation as they relate to this summarized data. We concentrate on Wetherill's approach since it is the foundation for our design strategy.

The design issues for these procedures are similar to those given in Section 2.2.1 for nonparametric procedures. Stimulus spacing and starting value potentially affect the informational content of the data collected. This is apparent in the approach of McLeish and Tosh [1983], where the approximate distribution used for estimation arises in the limit as the spacing width tends to zero and the initial design point tends to negative infinity. The Up and Down strategy, the design used in Wetherill's [1963] UDTR, also depends on stimulus level spacing. Einbinder [1973] suggested using Langlie's [1962] "One Shot Test Strategy" instead of the Up and Down method in an effort to diminish the potential for spacing problems. However, the advantage to this is suspect since Langlie's strategy has not been shown to be clearly superior.

Estimation is handled differently by these two methods. McLeish and Tosh [1983] used an MLE for which the original response distribution assumption is logistic. However, by restructuring the problem they were able to estimate its parameters using an exponential distribution. They showed empirically that their procedure makes better use of the collected data. The importance of the logistic assumption has not been addressed in detail though reasonable robustness to distributions proportional to $ce^{-\alpha x}$ is expected [McLeish and Tosh 1983]. Empirically, normal response distribution quantiles are estimated well with this approach. On the other hand, the UDTR strategy requires no distributional assumption since it estimates using the nonparametric w estimator. Since the quantile of interest is always the median quantile on the transformed response curve, the estimate should be reasonably robust. An important shortcoming of the UDTR is that it can provide only estimates of quantiles such as those listed in Table 2.1. Thus the experimenter is somewhat restricted in his ability to draw inference regarding the response distribution.

2.3 A New Approach to Extreme Quantile Estimation

In this section we introduce our new approach. We propose a specific technique in Section 2.3.1 which is intended to be a melding of work concerning estimation of the median to that of extreme quantile estimation. We include a preliminary justification for this technique, noting the suspected advantages according to the issues addressed in Section 2.2. Section 2.3.2 outlines the results pursued in this paper. Complete success cannot be claimed for each, but contributions to this area are made.

2.3.1 New Approach Application and Preliminary Justification

The transformed response curve of Section 2.1.3 is an attractive alternative to the original response function. It reduces the problem of extreme quantile estimation to the more practicable problem of estimation at or about the median. There, estimation is fairly robust to response function form. This robustness is essential given the lack of knowledge about the distribution tail. Additionally, Wetherill et al. [1966] showed the empirical performance to be good even when used in conjunction with a design and estimation technique which arguably could be improved upon. Thus, we propose to design and estimate in terms of the transformed response curve.

Among the available sequential design and estimation procedures, we feel that a hybrid strategy has much promise to be successful here. We intend to collect data with the Delayed Robbins-Monro (DRM) design and estimate quantiles of the transformed response curve with parametric maximum likelihood. This procedure has shown good empirical performance relative to several other common methods [Bodt and Tingey 1986]. It is robust to the selection of both initial design points and grossly inappropriate values of the constant c. The sequential nature of the design makes it resource-efficient. In terms of median estimation it is fairly robust to asymmetric response distribution forms--important since it is quite likely that the transformed response curve will be asymmetric. Additionally, using this method we can estimate quantiles other than those listed in Table 2.1. Computationally, with commonly available facilities, it is a feasible approach. Thus, we propose to use DRM as the design and parametric maximum likelihood as the estimation strategy.

The use of parametric maximum likelihood requires the selection of a parametric form to represent the response distribution. Certainly, if we do not know the form of the original response distribution we will not know the form of the transformed response distribution. Our approach is to select one which can assume many shapes. A three parameter model offered by Prentice [1976] for the original response distribution can be used here for the transformed response distribution. It is given by

$$T\{P(x)\} = \{1 + e^{-\frac{(x-\mu)}{\sigma}}\}^{-m} \tag{2.2}$$

Referring back to Table 2.1 note that many of the suggested transformations are of the form $T\{P(x)\} = P(x)^m$. In consideration of these transformations, if the common logistic assumption were valid, (2.2) exactly represents the form of the true $T\{P(x)\}$. This is an appealing feature. We refer to (2.2) *hereafter as the power logistic* distribution. Additionally, this distribution may assume asymmetric shapes depending on the value chosen for m, allowing the estimation procedure necessary flexibility. Thus, we propose to use the power logistic distribution for our parametric assumption.

2.3.2 Theoretical and Empirical Results Sought

In Chapters 3 and 4 we discuss the properties of the proposed scheme which will be hereafter referred to as the Power Logistic Transformed Response (PLTR) strategy. In Chapter 3 we address several points analytically. We show the development of $\hat{x}_{100p}$ for the PLTR and give its asymptotic distribution. Questions of estimate existence in finite samples and optimal design are also considered. In Chapter 4 we subject the scheme to a feasibility study in the form of a Monte Carlo exercise.

3. SOME ANALYTICAL RESULTS

The new approach to extreme quantile estimation, introduced in Section 2.3.1, joins three independent concepts: maximum likelihood estimation assuming a power logistic response distribution, the Delayed Robbins-Monro design, and the strategy of transformed responses. Each of these three possess their own set of desirable properties when brought to bear on problems of this type, but only those properties consonant with the combined approach goal of extreme quantile estimation are of interest. Since it is maximum likelihood which delivers the final quantile estimate, we must relegate DRM and transformed responses to strictly supportive roles where their properties are important only in contributing to the collection of good data. By good we mean data about the target quantile permitting maximum likelihood estimation. This chapter's structure reflects the belief that the analytical results are most important as they pertain to the final quantile estimate. Consequently, maximum likelihood serves as the structural focus, and DRM and transformed responses are addressed as they support maximum likelihood.

Chapter 3 consists of two sections. Section 3.1 develops maximum likelihood estimation for the power logistic distribution. Included in the development are the quantile estimators, their asymptotic properties, and some results regarding existence. Section 3.2 determines an optimal design and argues that the DRM strategy, acting on transformed responses, collects data in the spirit of optimality.

3.1. Maximum Likelihood Estimation Assuming the Power Logistic Distribution

In this section we detail the development of the MLEs for extreme quantiles. Recall that extreme quantiles for the true underlying response distribution are, through the strategy of transformed responses, quantiles about the median of the transformed response distribution. An assumption is that this transformed response distribution can be reasonably expressed in terms of the power logistic distribution suggested by Prentice [1976]. Thus MLEs for quantiles about the median of the power logistic distribution are used to estimate the desired quantiles of the underlying response distribution. In Section 3.1.1 we define the estimator and develop its asymptotic variance. In Section 3.1.2 we address estimate existence in finite samples.

3.1.1 Maximum Likelihood Estimator for x_{100p}

Let the transformed response distribution T(x) have the form

$$T(x) = \{e^y/(1 + e^y)\}^m, \tag{3.1}$$

for location $\mu \in (-\infty, \infty)$, scale $\sigma \in (0, \infty)$, and shape $m \in (0, \infty)$, where $y = (x - \mu)/\sigma \ \forall x \in (-\infty, \infty)$. Hereafter, we refer to the parameter space for m as Ω^1, for μ and σ together as Ω^2, and for all three together as Ω^3. We have dropped the notation T{P(x)} of Section 2.1.3 for this section because it emphasizes the mathematical transformation of responses. Here it is important to emphasize that the transformed response distribution has an assumed form, given by (3.1), and is not simply the range of a transformation.

The likelihood function arises as follows. Let s_i and b_i represent the number of successes and trial blocks, respectively, at stimulus x_i. Recall that the strategy of transformed responses requires samples to be taken block sequentially until a success or failure is observed. The probability that a success occurs at stimulus x_i is the probability x_i exceeds the tolerance for each subject tested in this trial block, that is, each subject responded to the stimulus. This probability is modeled by the transformed response distribution and is given by $T(x_i)$, denoted simply T_i. Thus, assuming subject independence, the probability of s_i successes in b_i trial blocks follows the binomial distribution with success probability T_i. It follows that the consideration of k levels of stimulus leads to the likelihood function expressed as

$$L(\underline{s};\underline{\theta}) = \prod_{i=1}^{k} \binom{b_i}{s_i} T_i^{s_i} (1 - T_i)^{b_i - s_i}, \tag{3.2}$$

where $\underline{s} = \{s_i; i = 1,2,...,k\}$, and $\underline{\theta} = \{\theta_1 = \mu, \theta_2 = \sigma, \theta_3 = m\}$, the parameters on which the success probabilities T_i depend.

The parameter values which maximize (3.2) can theoretically be determined in a straightforward way through the use of the log-likelihood. The log-likelihood, denoted l, corresponding to (3.2) is given by

$$l = \log L(\underline{s};\underline{\theta}) = \sum_{i=1}^{k} \{\log \binom{b_i}{s_i} + s_i \log (T_i) + (b_i - s_i) \log(1 - T_i)\}.$$

The first derivative of the log-likelihood with respect to θ_j is given by

$$\frac{\partial l}{\partial \theta_j} = \sum_{i=1}^{k} \{s_i/T_i - (b_i - s_i)/(1 - T_i)\} \{\partial T_i/\partial \theta_j\}. \tag{3.3}$$

Solution of the system of equations $\{\frac{\partial l}{\partial \theta_1}, \frac{\partial l}{\partial \theta_2}, \frac{\partial l}{\partial \theta_3}\}^T = \underline{0}$ with respect to $\underline{\theta}$ yields the desired estimates.

It is useful to pause at this point to define two notational conveniences. First, we define a function $G(x) = e^y/(1 + e^y)$. Note that $G(x)$ has the form of a logistic distribution. Second, each of G, T, and y is a function of the stimulus x, but in what follows the x is omitted in the notation. When a specific i^{th} stimulus, x_i, is indicated, the subscript i will accompany G, T, and y.

The first deviatives of l with respect to each of the θ_j's differ only by $\partial T_i/\partial\theta_j$. The representations of $\partial T_i/\partial\theta_j$ for each parameter μ, σ, and m are now developed for their inclusion in (3.3). The derivative of G appears as

$$dG/dy = [e^y/(1 + e^y)]\,[1/(1 + e^y)] = G(1 - G) \tag{3.4}$$

Then remembering $y = (x - \mu)/\sigma$ it follows from (3.4) that

$$\partial G/\partial\mu = G(1 - G)\,\partial y/\partial\mu = G(1 - G)\,(-1/\sigma), \tag{3.5}$$

$$\partial G/\partial\sigma = G(1 - G)\,\partial y/\partial\sigma = G(1 - G)\,(-y/\sigma). \tag{3.6}$$

Noting that $T = G^m$ and using (3.5) and (3.6) we have

$$\partial T/\partial\mu = m\,G^{m-1}\,\partial G/\partial\mu = m\,T(1 - G)\,(-\,1/\sigma), \tag{3.7}$$

$$\partial T/\partial\sigma = m\,G^{m-1}\,\partial G/\partial\sigma = m\,T(1 - G)\,(-y/\sigma), \tag{3.8}$$

$$\partial T/\partial m = T\log(G). \tag{3.9}$$

The explicit form of (3.3) for each parameter can now be given. Combining (3.3) separately with each of (3.7) - (3.9) yields the appropriate derivatives

$$\partial l/\partial \mu = \sum_{i=1}^{k} \left\{ \frac{s_i - b_i T_i}{T_i (1 - T_i)} \right\} m\, T_i (1 - G_i) (-1/\sigma) \tag{3.10}$$

$$\partial l/\partial \sigma = \sum_{i=1}^{k} \left\{ \frac{s_i - b_i T_i}{T_i (1 - T_i)} \right\} m\, T_i (1 - G_i) (-y_i/\sigma) \tag{3.11}$$

$$\partial l/\partial m = \sum_{i=1}^{k} \left\{ \frac{s_i - b_i T_i}{T_i (1 - T_i)} \right\} T_i \log G_i\,. \tag{3.12}$$

The solution of the system formed by setting (3.10) - (3.12) equal to 0 yields the MLEs $\hat{\mu}$, $\hat{\sigma}$, and $\hat{m}$. The numerical approach used to deliver the estimators is discussed in Chapter 4.

The p^{th} quantile, x_{100p}, may be estimated using the above results. We simply solve (3.1) for the stimulus x for which $T(x) = q$. Recall from Section 2.1.3 that the stimulus corresponding to the q^{th} quantile of $T(\cdot)$ is the p^{th} quantile of $P(x)$, the tolerance distribution. Then we may write the MLE for x_{100p} as

$$\hat{x}_{100p} = \hat{\mu} + \hat{\sigma}\, \{-\log (q^{-1/\hat{m}} - 1)\} \tag{3.13}$$

by the invariance property of maximum likelihood.

The asymptotic properties of $\hat{x}_{100p}$ follow from maximum likelihood. The estimator $\hat{x}_{100p}$ is consistent, efficient, and normally distributed. The asymptotic variance of $\hat{x}_{100p}$ may be approximated from a truncated Taylor series expansion of $\hat{x}_{100p}$ about (μ, σ, m), yielding

$$\mathrm{Var}\,(\hat{x}_{100p}) \approx \hat{c}^T\, I^{-1}\, \hat{c},$$

where $\hat{c}$ is the gradient vector of $\hat{x}_{100p}$ and I is Fisher's Information. The gradient vector is easy to compute and is given by

$$\hat{c}^T = \{1, -\log(q^{-1/\hat{m}} - 1), -\hat{\partial}\,\hat{m}^{-2}\log(q)/(1 - q^{-1/\hat{m}}\}.$$

Remaining is an expression for the information matrix I.

The information matrix is given by $I_{jh} = E(\frac{\partial l}{\partial\theta_j} \cdot \frac{\partial l}{\partial\theta_h})$. The product of the two first derivatives is expressed by

$$\frac{\partial l}{\partial\theta_j} \cdot \frac{\partial l}{\partial\theta_h} = \sum_{i=1}^{k} \left\{\frac{s_i - b_i T_i}{T_i(1-T_i)}\right\}^2 \{\partial T_i/\partial\theta_j\}\{\partial T_i/\partial\theta_h\} +$$

$$\sum_{i}^{k} \sum_{\neq r}^{k} \left\{\frac{s_i - b_i T_i}{T_i(1-T_i)}\right\}\left\{\frac{s_r - b_r T_r}{T_r(1-T_r)}\right\} \{\partial T_i/\partial\theta_j\}\{\partial T_r/\partial\theta_h\}. \qquad (3.14)$$

I_{jh} is then the sum of the expectation of each of these two expressions. Consider the second one first. Since each trial, and hence each trial block, is considered independent, the random variables s_i and s_r for $i \neq r$ are independent also. Note that the random variable s_i is binomial with expectation b_iT_i. Then

$$E\left\{\frac{s_i - b_i T_i}{T_i(1-T_i)}\right\} = \left\{\frac{b_i T_i - b_i T_i}{T_i(1-T_i)}\right\} = 0,$$

and by independence the expectation of the second term of (3.14) is zero. Thus I_{jh} consists only of the expectation of the first term. Noting that

$$E(s_i^2) = b_i T_i (1-T_i) + b_i^2 T_i^2$$

we can compute

$$E\{s_i - b_i T_i\}^2 = E\{b_i T_i (1-T_i) + b_i^2 T_i^2 - 2 b_i^2 T_i^2 + b_i^2 T_i^2\} = b_i T_i (1-T_i).$$

Hence,

$$I_{jh} = \sum_{i=1}^{k} \{b_i/T_i (1-T_i)\} \{\partial T_i/\partial \theta_j\} \{\partial T_i/\partial \theta_h\}, \tag{3.15}$$

the last two terms coming from (3.7) - (3.9).

3.1.2. Conditions for the Existence of MLEs in Finite Samples

In this section we establish conditions on the data guaranteeing the existence of bounded MLEs of some of the parameters (μ, σ, m). It is well known that in binomial response models the estimates, though asymptotically existing with probability tending to 1, may not exist for certain finite samples. See for example Wedderburn [1976] or Silvapulle [1981]. The power logistic form used here also may not admit a solution for a particular sample. An argument built on convexity delivers the data restrictions needed for estimation to be possible.

We begin by citing the theorem of Silvapulle [1981]. Define $F(\alpha + \eta x)$ as a distribution function dependent on the linear parameters (α, η), and assume that there are at least two distinct stimulus levels. The MLE of (α, η) is denoted ($\hat{\alpha}$, $\hat{\eta}$). For each response type, form ordered levels of the stimulus, and denote them by $x_{(1)}^j, x_{(2)}^j, \cdots, x_{(n)}^j$, where j = 0 or 1 according to the observance of a nonresponse or response, respectively. So as not to confuse response type with stimulus level order, we replace the subscript, (1), with (min) and the subscript, (n), with (max). From Wu [1985], the condition Π is defined as responses and nonresponses occurring on the stimulus axis in one of the following three ways.

1. If $x^1_{(min)} \neq x^1_{(max)}$ and $x^0_{(min)} \neq x^0_{(max)}$,
   then $(x^1_{(min)}, x^1_{(max)}) \cap (x^0_{(min)}, x^0_{(max)}) \neq \emptyset$.

2. If $x^0_{(min)} = x^0_{(max)} = x^*$,
   then $x^1_{(min)} < x^* < x^1_{(max)}$.

3. If $x^1_{(min)} = x^1_{(max)} = x^{**}$,

then $x^0_{min} < x^{**} < x^0_{max}$.

The first possibility is what engineers refer to as a zone of "mixed" results, an interval on the stimulus axis which contains both responses and nonresponses. The latter two address the situations in which a single stimulus induces all of the observed responses or all of the observed nonresponses. For the special case of the two-parameter distribution, F, Theorem iii from Silvapulle [1981] may be stated:

Theorem 3.1 Suppose that - log (F) and - log (1-F) are convex. Then $(\hat{\alpha}, \hat{\eta})$ exists and the minimum set $\{(\hat{\alpha}, \hat{\eta})\}$ is bounded if and only if Π is satisfied. Let us further assume that F is strictly increasing at every t satisfying $0 < F(t) < 1$. Then $(\hat{\alpha}, \hat{\eta})$ is uniquely defined if and only if Π is satisfied.

The notion of a "minimum set" is discussed by Silvapulle [1981]. The boundedness is with respect to the minimization of the negative likelihood, $- l(\hat{\alpha}, \hat{\eta})$.

The power logistic distribution (3.1) which we have selected to model the transformed responses has an additional parameter, m. However, for any fixed m in Ω^1 it will be shown that bounded, unique MLEs $(\hat{\mu}, \hat{\sigma})$ exist for (μ, σ) if and only if condition Π is satisfied.

Theorem 3.2 Assume $m \in \Omega^1$ is known. Then Π is both necessary and sufficient for the bounded unique existence of $(\hat{\mu}, \hat{\sigma})$, the MLEs for the remaining parameters (μ, σ) of T.

Proof: The proof is an application of Theorem 3.1 to T. We claim that - log(T) and - log(1-T) are convex, and we show this as Property 3.1. To show that T is strictly increasing, consider that $G \in (0,1)$, $\forall (\mu, \sigma, m) \in \Omega^3$ and $x \in (-\infty, \infty)$. It follows that $\frac{dT}{dx} = m\, G^m (1-G)\, 1/\sigma > 0$. Finally, the parameters of T can be expressed as $\alpha + \eta x$ by setting $\alpha = -\mu/\sigma$, $\eta = 1/\sigma$. Application of Theorem 3.1 establishes the result for (α, η), and from the 1-1 correspondence between (α, η) and (μ, σ) the assertion follows. ■

Property 3.1 Both - log(T) and - log(1-T) are strictly convex.

Proof: Since both functions are twice differentiable, it is sufficient to show that their second derivatives are positive. Without loss of generality we will differentiate with respect to y instead of x to simplify the expressions. The chain rule and positivity of σ allow this simplification.

Proof for - log(T) To begin, $\frac{d}{dy}[-\log(T)] = \frac{-m\, G^{m-1}\, G\, (1-G)}{G^m} = m\,(G-1)$. Then $\frac{d^2}{dy^2}[-\log(T)] = m\, G\, (1-G) > 0$. ■

Proof for -log(1 - T) To begin,

$$\frac{d}{dy}[-\log(1-T)] = \frac{-[-m\,G^{m-1}\,G\,(1-G)]}{1-G^m} = \frac{m\,G^m\,(1-G)}{1-G^m}.$$ Let

$w(y) = \dfrac{m\,G^m\,(1-G)}{1-G^m}$. Since $w(y) > 0\ \forall y \in (-\infty, \infty)$ it follows that $\dfrac{dw}{dy} > 0$ iff $\dfrac{d}{dy}[\log w(y)] > 0$. Then,

$\log w(y) = \log(m) + m\log(G) + \log(1-G) - \log(1-G^m)$ and

$\dfrac{d}{dy}[\log w(y)] = m\,\dfrac{dG/dy}{G} - \dfrac{dG/dy}{1-G} + \dfrac{mG^{m-1}\,dG/dy}{1-G^m}$. Since $\dfrac{dG}{dy}$, the logistic density, is positive and appears in each term it is sufficient to show that

$\dfrac{m}{G} - \dfrac{1}{1-G} + \dfrac{mG^{m-1}}{1-G^m} > 0$. Combining the fractions yields a denominator $G\,(1-G)\,(1-G^m) > 0$ and a numerator $N = m\,(1-G) - G\,(1-G^m)$. It remains only to show that $N > 0$. Observe that $\lim_{G\downarrow 0} N = m$; $\lim_{G\uparrow 1} N = 0$. Finally consider $\dfrac{dN}{dG}$ which after simplification can be expressed $(m+1)\,(G^m - 1)\,\dfrac{dG}{dy} < 0$, and is strictly decreasing. ∎

Theorem 3.2 establishes necessary and sufficient conditions on the data for bounded and unique MLEs of (α, η) providing $m \in \Omega^1$ is known. In what follows, we establish the necessary and sufficient condition for the unique existence of an MLE for m, with $(\mu, \sigma) \in \Omega^2$ known. The derivative of the log-likelihood with respect to m appears as (3.12). In this expression, for our application to transformed responses, $b_i = 1, \forall i$, and $s_i = 0$ or 1 corresponding to a failure or success respectively. Then substituting $b_i = 1$ and recalling that $T_i = G_i^m$, we recognize that a stationary point will occur with respect to m when

$$dl/dm = \sum_i \{s_i - T_i\}\,\frac{\log G_i}{1-G_i^m} = 0. \tag{3.16}$$

The following theorem identifies a necessary and sufficient condition for the data under which a stationary point is present and is a unique maximum.

Theorem 3.3 Assume $(\mu, \sigma) \in \Omega^2$ are known. Then the presence of at least one success and one failure is necessary and sufficient for the unique existence of a maximum likelihood estimator of m.

Proof: Necessity is proved first. Note that $\dfrac{\log G_i}{1-G_i^m}$ is strictly negative,

$$s_i - T_i = 1 - T_i > 0 \text{ for } s_i = 1 \text{ (success),}$$

and

$$= \ - T_i < 0 \text{ for } s_i = 0 \text{ (failure).}$$

The i^{th} summand of (3.16) is negative if the i^{th} stimulus block results in a success, and otherwise is positive. Then if (3.16) is true, it is necessary that at least one success and one failure be observed. That the presence of at least one success and one failure is sufficient is proven by showing a unique solution to (3.16) exists and represents a maximum. Consider the case in which both successes and failures have occurred. Failures are indexed over i, successes over j. Then solve

$$\sum_{i,j} \{s_k - T_k\} \frac{\log G_k}{(1-G_k^m)} = 0.$$

It follows that

$$\sum_i -T_i \frac{\log G_i}{(1-G_i^m)} = -\sum_j (1-T_j) \frac{\log G_j}{(1-G_j^m)}$$

and

$$\sum_i \log G_i \frac{G_i^m}{(1-G_i^m)} = \sum_j \log G_j. \tag{3.17}$$

Note that the left hand side (LHS) satisfies $\lim_{m\downarrow 0} \text{LHS} = -\infty$ and $\lim_{m\uparrow\infty} \text{LHS}=0$. Further

$$\frac{d}{dm}\,\text{LHS} = \sum_i \log G_i \, \frac{(1-G_i^m)\, G_i^m \log G_i - G_i^m(-\,G_i^m \log G_i)}{(1-G_i^m)^2}$$

$$= \sum_i (\log G_i)^2 \frac{G_i^m}{(1-G_i^m)^2} > 0.$$

Therefore, since $\sum \log G_j < 0$ and LHS is strictly increasing there exists a unique m for which the log-likelihood equation is solved. That the solution must be a maximum under the data conditions comes directly from the second derivative.

$$\frac{d^2}{dm^2} = \frac{d}{dm} \sum (s_k - T_k) \frac{\log G_k}{(1-G_k^m)}$$

$$= \sum \{(s_k - T_k) \frac{(1-G_k^m)(0) - (\log G_k)(-G_k^m \log G_k)}{(1-G_k^m)^2}$$

$$+ \frac{\log G_k}{(1-G_k^m)}(-\,G_k^m \log G_k)\}$$

$$= \sum \left\{ \frac{(\log G_k)^2\, G_k^m}{(1-G_k^m)^2} \right\} (s_k - 1).$$

The quantity { } is always positive. Since $s_k = 0$ or 1, d^2/dm^2 will always be negative so long as at least one failure is observed. Any stationary point must be a local maximum. ■

The above does not preclude unbounded m. In Theorem 3.4 we impose some additional conditions in order to provide a bounded solution for m.

Theorem 3.4 Suppose

(a) there exists bounds L, U such that $\forall i, j$, $0 < L < \left(\frac{G_i}{G_j}\right) < U < 1$,

(b) there exists m such that $0 < m < \infty$. Claim: If the parameter m and the finite collection of probabilities $(G_i)_i$ and $(G_j)_j$ satisfy

$$\text{(t)} \quad \sum_i \log G_i \frac{G_i^m}{1 - G_i^m} = \sum_j \log G_j,$$

then there exists bounds m_0 and m_1 such that $0 < m_0 < m < m_1 < \infty$.

Proof: It follows from (a) that each of $\log G_i$ and $\log G_j$ is bounded and bounded away from 0. Further, the finite sum over j of $\log G_j$, the RHS of (t), is bounded and bounded away from 0. It follows from (a) and (b) that $\frac{G_i^m}{1-G_i^m} > 0 \; \forall G_i, m$. Let $<(G_i^{(\alpha)})_i>_\alpha$, $<(G_j^{(\alpha)})_j>_\alpha$ denote sequences of $(G_i)_i$ and $(G_j)_j$, respectively, indexed over α. Suppose there exists $<(G_i^{(\alpha)})_i>_\alpha$, $<(G_j^{(\alpha)})_j>_\alpha$, and corresponding $<m^\alpha>$ satisfying (t), and in addition either (1) $<m^\alpha> \to \infty$ or (2) $<m^\alpha> \to 0$. *Case (1):* If $<m^\alpha> \to \infty$, then $\frac{G_i^m}{1-G_i^m} \to 0$, and thus LHS $\to 0$. Therefore, under (t) the RHS $\to 0$, which contradicts the RHS being bounded away from 0. Therefore (t) cannot be satisfied and $<m^\alpha> \to \infty$. This implies that $<m^\alpha>$ is bounded, say $m < m_1 < \infty$. (*)

Case (2): If $<m^\alpha> \to 0$, then $\frac{G_i^m}{1-G_i^m} \to \infty$. Thus LHS $\to -\infty$ since $\log G_i < 0$. Therefore, under (t) the RHS $\to -\infty$ which contradicts the RHS being bounded. Therefore (t) cannot be satisfied and $<m^\alpha> \to 0$. This implies that $<m^\alpha>$ is bounded away from 0, say $0 < m_0 < m$. (**)

Combining (*) and (**), $0 < m_0 < m < m_1 < \infty$. ∎

3.2. Robbins-Monro and Optimal Design for Estimating x_{100p}

In this section we demonstrate the appropriateness, in application to this problem, of the Stochastic Approximation Method of Robbins and Monro [1951]. Recall that in our procedure, the Robbins-Monro strategy is used only as a sequential design, supportive to maximum likelihood estimation of the target quantile. In order to show its usefulness, we have to discuss optimal design in this problem's context, and then show that the Robbins-Monro strategy approximates the optimal design in some sense. Moreover, an argument is made for its suitability in collecting data likely to satisfy condition II of the previous section. This section consists of two parts. Section 3.2.1 discusses the issue of optimal design, and Section 3.2.2 argues that the Robbins-Monro strategy acts both in the spirit of optimality and in deference to practical concerns.

3.2.1. Optimal Design for Estimating x_{100p}

There are many different approaches to optimality. See, for example, Chernoff [1979], Fedorov [1972], or Silvey [1984]. To begin we must select one well suited to our estimation goal. Our principal concern is that data be collected to support the power logistic maximum likelihood estimation of a specific quantile. With the variance of the estimate as our measure of closeness, a reasonable approach is to distribute available samples over stimulus levels so as to minimize the variance of the estimate for each possible target quantile x_{100p}. An approximate asymptotic procedure which approaches optimality in this way is referred to as c-optimality by Silvey [1980]. This section discusses a c-optimal approach presented by Wu [1987] and its application to the power logistic maximum likelihood estimation of x_{100p}. For this discussion we assume that m is known. Consider the estimate for x_{100p} to be given by the function

$$g(\mu, \sigma) = \hat{\mu} + \hat{\sigma}\{-\log(q^{-1/m} - 1)\}.$$

The elements of the information matrix $I(\mu, \sigma)$ may be taken directly from the corresponding elements of the information matrix given by (3.15). Then proceeding as in Section 3.1.2, the asymptotic variance of $G(\hat{\mu}, \hat{\sigma})$ is locally approximated by

$$\nabla g^T I^{-1}(\mu, \sigma) \nabla g. \tag{3.18}$$

Let $b = \sum b_i$, the total number of observations in all blocks. Normalizing $I(\mu, \sigma)$ by dividing each b_i by b provides a solution which is independent of the sample. The c-optimal design is then obtained by minimizing, after normalization, (3.18) with respect to the choices of the stimulus levels x_i and $\lambda_i = \frac{b_i}{b}$, where $\lambda_i > 0, \sum_1^k \lambda_i = 1$. The inverse of I is generalized and permitted to be singular since optimal solutions to the design problem often consist of allocating all resources at one stimulus level.

The general solution to this optimization problem is given by Wu [1987] and applied to the power logistic distribution. After presenting the general results we need to examine optimal designs for specific values of m in greater detail. Define

$$w(y) = \frac{dT(y)}{dy} / \{T(y)\,(1 - T(y)\}^{1/2}, \tag{3.19}$$

and let the slope of the curve $c = w(y)\,(1, y)$ be given by

$$r(y) = \frac{d}{dy}\,[w(y)\,y] \Big/ \frac{d}{dy}\,w(y) = y + w(y) \Big/ \frac{d}{dy} w(y).$$

Then the relevant portion of the theorem given by Wu [1987] drawn from Ford, Tornsey and Wu [1988] is listed as Theorem 3.5.

Theorem 3.5 Suppose that the curve c is closed, bounded and convex. Then there exists $p_1 = T(\underline{y})$ and $p_2 = T(\overline{y})$, $p_1 < p_2$, with $\underline{y}$ and $\overline{y}$ satisfying (3.20),

$$r(\overline{y}) = r(\underline{y}) = \frac{w(\overline{y})\,\overline{y} + w(\underline{y})\underline{y}}{w(\overline{y}) + w(\underline{y})} \tag{3.20}$$

such that if $p_1 \le p \le p_2$, then the c-optimal design for estimating x_p is the one-point design allocating all samples to the stimulus level corresponding to p.

The conditions of the theorem hold for the power logistic distribution [Wu 1987]. Some values for p_1 and p_2, taken from Wu [1987], are listed in Table 3.1.

Table 3.1 Range of quantiles calling for a one-point c-optimal design

m	range $[p_1, p_2]$
1/2	[.110, .939]
2/3	[.097, .925]
3/2	[.069, .909]
3	[.051, .898]

Note for the few values of m considered, $p_1 < .5 < p_2$ which, from Theorem 3.5, indicates that the c-optimal design for x_{50} calls for sample allocation of the stimulus level corresponding to the median. We now extend these results.

Recall that the PLTR strategy transforms responses in such a way that the target quantile is local to the median of the assumed power logistic distribution. We assert that in all practical cases, the c-optimal design for estimating this target quantile is the one-point design placing all observations at the target quantile. By practical cases, we mean those strategies suggested in Table 2.1, where $n_c \approx m$ ranges from two to fourteen. (One would expect them to be exactly equal only under large sample conditions and under the assumption that the original response function family was known to be logistic.) We addressed the assertion numerically by finding $p_1 < p_2$ for which (3.20) was satisfied. The values $p_1 < p_2$ were chosen on (0, 1) in increments of .00001 for each value of m on [.1, 20] chosen in increments of .1. The values $\underline{y}$ and $\overline{y}$ were computed by $T^{-1}(p_1)$ and $T^{-1}(p_2)$, respectively, and then inserted in (3.20) to check for equivalence. The resulting p_1 and p_2 appear graphically in Figure 3.1 and a subset appears in Table 3.2. Numerically, the assertion of c-optimality holds, as not only do the curves for p_1 and p_2 not cross the .5 quantile, but also they appear to be approaching asymptotes far removed from .5; over the interval $m \in [10.2, 20]$, p_1 decreases only by .002 and p_2 by .003.

Table 3.2 Range of quantiles calling for a one-point c-optimal design (extended)

m	range $[p_1, p_2]$
.1	[.124, .962]
.3	[.111, .943]
.5	[.105, .933]
.7	[.096, .926]
.9	[.087, .921]
1.1	[.080, .917]
1.3	[.074, .914]
1.5	[.069, .911]
1.7	[.065, .909]
1.9	[.062, .907]
2.0	[.061, .906]
4.0	[.046, .897]
8.0	[.037, .891]
12.0	[.034, .889]
16.0	[.033, .888]
20.0	[.032, .887]

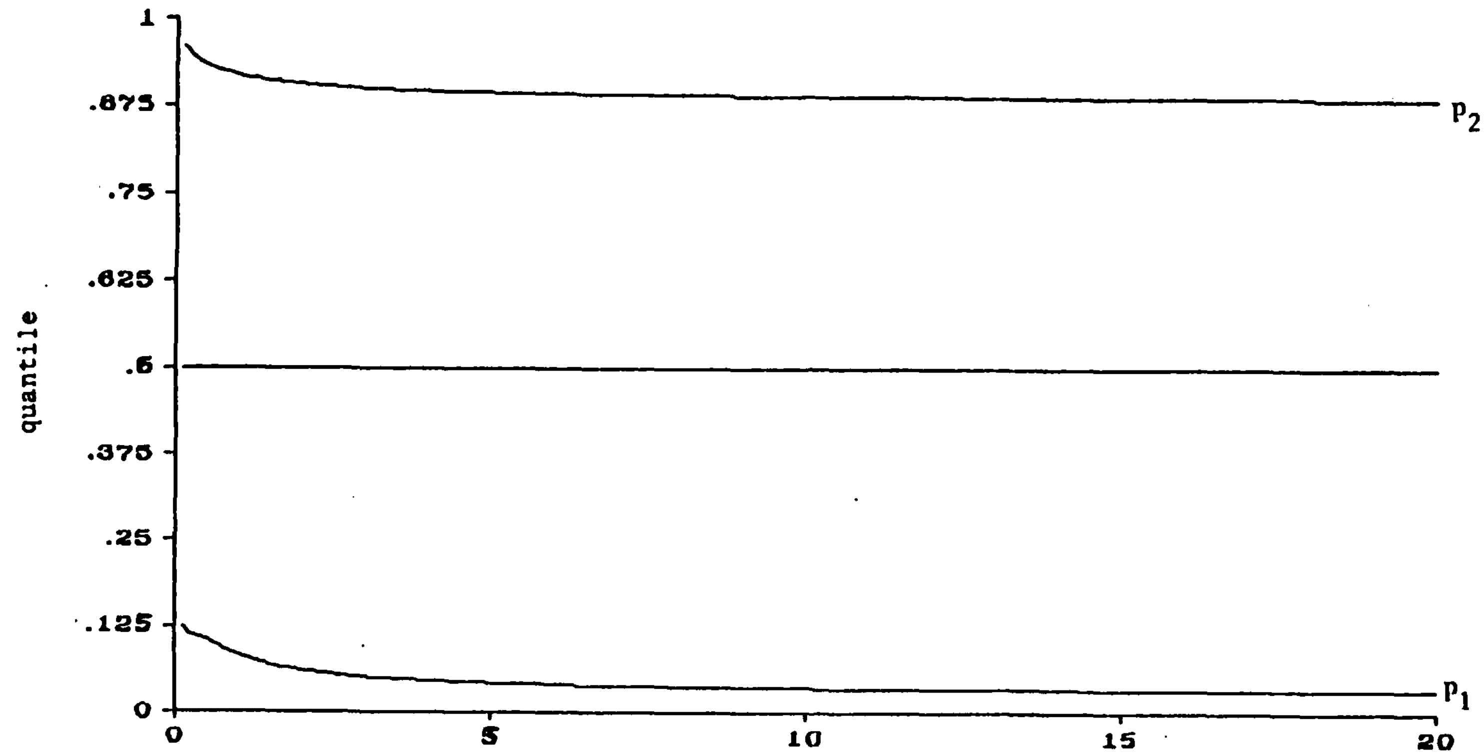

Figure 3.1 Range of p_1 and p_2 calling for a one-point c-optimal design taken over $m \in$ (.[illegible]0,20)

3.2.2. The Appropriateness of Robbins-Monro

In this section we argue that the Robbins-Monro strategy, specifically the Delayed Robbins-Monro which we are employing, has desirable properties with respect to the c-optimality and the existence of MLEs. Beginning with c-optimality, recognize that there is no way of truly attaining a c-optimal design for this problem. Its complete determination requires prior knowledge of the target quantile location, a location dependent on unknown parameters. In two stage testing one could use a first stage to estimate the parameters and then allocate samples based on those estimates. We have chosen instead to approach the problem sequentially and without reference, during data collection, to a specific parametric family. Two important properties support that choice.

The original Robbins and Monro [1951] strategy is given by

$$x_{n+1} = x_n - a_n (y_n - p),$$

where x_n is the n^{th} design point, y_n is 0 or 1, according as failure or success is observed, and a_n is a decreasing sequence of numbers tending to 0. Under very general conditions, Dvoretzky [1956] showed that the sequence of design points converges in probability to p. Then in terms of large sample sizes, this procedure is gathering data in the spirit of c-optimality. Unfortunately, we do not have very large sample sizes at all. It is then very important for the process to converge rapidly. What the experimenter can alter is the choice of the sequence, represented by a_n. Chung [1954] establishes that setting $a_n = c/n$ causes rapid convergence to the target quantile. Kesten [1958] extended this result by showing this applies also to variations in which the magnitude of the difference between x_n and x_{n+1} is dependent on the number of reversals, as in the case for DRM. The DRM then is a procedure which converges rapidly in probability to the target quantile. Therefore, theoretically, it is well suited to gather information approximating a c-optimal design. In application, even for quite small samples, it has been shown to perform well in this task. Several of the articles cited in Chapter 1 support the small sample performance claim.

With regard to estimate existence, we argue informally that the DRM collects data in such a way as to maximize the probability the MLEs for (μ, σ) exist. Recall from Theorem 3.2 that condition Π is sufficient to ensure MLE existence. An essential aspect of Π is that there exist an $x_1 < x_2$ where x_1 results in a success and x_2 a failure. The probability that this occurs can be expressed $p(s_1 = 1)\, p(s_2 = 0)$ by trial independence. Suppose the goal is to choose x_1 and x_2 with $x_1 < x_2$ so as to maximize that probability. The fairly obvious solution, subject to the inequality, is to select x_1 and x_2 so that $p(s_1 = 1) \approx p(s_2 = 0) \approx 1/2$. Then our chances of satisfying Π are greatly enhanced by taking many observations close to the median of the transformed response distribution. From the previous discussion, the DRM is appropriate for collecting such observations. So not only does the DRM strategy collect data in the spirit of c-optimality, but by virtue of the transformed response median being the target, it also collects data in an efficient manner in terms of estimate existence.

4. A SIMULATION STUDY

Simulation provides an arena in which the PLTR strategy may be studied in operation. Through simulation we may explore the combined performance of the three distinct concepts that comprise our strategy. The Delayed Robbins-Monro design, the transformed response approach, and the maximum likelihood estimation of power logistic quantiles each has solid analytical footings supporting their individual use. However, their individual strengths do not guarantee the effectiveness of their use in combination. Moreover, those supporting analytical results are primarily asymptotic findings, lending only limited insight to practical applications. It is necessary to evaluate the procedural properties of this method. Simulation serves well in that task.

Some specific benefits of simulation follow. Through simulation we gain experience with the performance of the procedure in completely specified environments. We may determine its feasibility by considering issues such as estimate bias and small sample variance. Procedural quirks such as problems in convergence can be identified. In this chapter, we pursue each of these subjects in an attempt to better understand the method's properties.

Chapter 4 considers a simulation study of the PLTR strategy for the estimation of extreme quantiles. Section 4.1 discusses the scope of this exercise. Section 4.2 describes the simulation design and methods of analysis. Section 4.3 details the optimization method used. Section 4.4 presents the results and addresses the above issues.

4.1 Scope

The scope of this study is intentionally narrow. We seek only to accomplish two tasks. The first is to determine the feasibility of applying this strategy to extreme quantile estimation. We leave detailed performance evaluation for further study. The second is to relate, where appropriate, the analytical results of Chapter 3 to the empirical findings of this chapter. We hope to derive support for the strategy's concept from the simulation.

To address feasibility, we must determine whether or not the procedure works, and at what cost. We focus on the performance of the strategy for a small set of design conditions. The properties of most interest are bias, mean square error, and robustness to parametric form. Additional properties, such as estimate convergence, are considered to help gauge the cost of implementation.

Empirical support for the procedure is also pursued. Certainly, where empirical and analytical results can be compared, they must agree. But also they can combine to provide a more complete understanding of the strategy. Trivially, small sample behavior would be extremely difficult to study from a theoretical standpoint as would asymptotic behavior from an empirical one. Less obvious distinctions occur when considering, estimate existence, and optimal design. Each have rigid analytical interpretations, but also they have practical implications. Our intent is to approach these issues in both manners, drawing upon pertinent analytical results from Chapter 3 and adding to them the empirical results of this chapter.

4.2 Experimental Strategy

In this section we discuss the experimental strategy for the simulation. Subsections are termed simply Design and Analysis. In Section 4.2.1 we introduce the design, justify its components, and describe how the data is to be collected. In Section 4.2.2 we summarize the intended analysis, noting its relationship to the objectives of this chapter.

4.2.1 Design

The design matrix is given in Table 4.1. It consists of just three factors: target quantile, distribution, and sample size. Target quantiles determine both the quantile to be estimated and the specific transformed response strategy chosen, that is, how many subjects must respond in a stimulus block before a success occurs. The distribution refers to the true underlying response distribution for which the target quantiles are to be estimated. The sample size indicates the number of stimulus blocks to be sampled. The primary response is the root mean square error $\sqrt{\text{mse}}$.

The .8, .9, and .95 quantiles serve as the levels of the target quantile factor. We are taking extreme quantiles to mean those outside the first and third quartiles but consider here only some commonly sought upper tail quantiles. For the two symmetric distributions there is no loss in generality by considering only the upper tail. These specific quantiles are not among the resultant median quantiles from the transformed response strategies listed in Table 2.1. However, they do closely approximate the transformed medians .7937, .9057, and .9517 suggested by the strategies requiring 3, 7, and 14 responses, respectively, for a success.

An interest in these three relates to their corresponding quantiles of the transformed response distribution. Optimal design results noted in Chapter 2 suggest that for several response distributions one should collect data precisely at the stimulus level corresponding to the quantile to be estimated. However, data intended for median estimation is commonly used to draw inference about the location of most quantiles between the first and third quartiles. Beyond them one runs greater risks of departure from the parametric assumption as well as extrapolated estimates. In our study, when the target quantile is .8 (a transformed quantile of .5) the corresponding transformed quantile of .9 is less than .75. Thus if we accept the above mentioned

Table 4.1 Design matrix

Sample Blocks	Response Distribution	Target Quantile .8	.9	.95
15	Cauchy	$\sqrt{mse}$		
	Exponential			
	Logistic			
20	Cauchy			
	Exponential			
	Logistic			

practice, we may estimate the .9 quantile using data which actually targets the .8 quantile. A similar relationship holds between .9 and .95 and also .95 and .99. If the ".75" estimate proves to be good, the resource advantage could be substantial. Recall, for example, that only 3 responses are required to establish a success when estimating the .8 quantile in contrast to 7 responses for the .9 quantile.

The distribution factor consists of three levels: the logistic, two-parameter exponential, and Cauchy distributions. The logistic is one of the most common response distributions assumed because of its similarity to the normal distribution and its greater mathematical tractability. In this study it has location and scale parameters of 0 and 1 respectively. The two-parameter exponential is the representative asymmetric response function. It has median 0 and variance equivalent to that of the logistic. The Cauchy distribution, with its heavy tails, is an obvious severe test case for any extreme quantile estimation. It has been scaled so that its first and third quartiles match those of the logistic. When Wetherill et al. [1966] investigated the use of transformed response strategies on Cauchy quantiles, the results were discouraging. Thus the Cauchy response distribution should pose a challenge for the proposed estimation method.

Two sample sizes, 15 and 20 are considered. These two were chosen for the reasons that follow. Previous experience [Bodt and Tingey 1986] suggests that the precision resulting from 15 samples would be acceptable here. The practical concern for too many required subjects bounds the sample size above at approximately 20 in the most resource exhaustive case. There, to target the .95 quantile, as many as 20 blocks x 14 subjects/block yielding 280 subjects could be required. This requirement, 280 subjects, is consistent with applications in the literature, but many more would be impractical.

Extreme quantile estimates, used in the computation of the response for the design matrix, are gathered as follows. The simulation chooses a treatment combination and produces 500 estimates for each quantile. We chose 500 iterations because of the acceptance of that number in the literature. For each estimate, data is gathered by the Delayed Robbins-Monro strategy acting on transformed responses. Estimation is based on the power logistic assumption and uses transformed responses. The only exception to this is when the distribution is logistic. Then maximum likelihood using the original responses and the logistic assumption is also employed. Those cases in which the response distribution is truly logistic afford us an opportunity to empirically address whether a disadvantage in estimation results when using the summarized transformed responses instead of the original responses.

Two issues remain to be discussed. One objective of this study is to compare the PLTR's performance for different response distributions. What we would like to isolate as the major cause for observed differences is the differently weighted tails. In an attempt to reduce confounding we tried to make them comparable with regard to spread. Recall that the logistic and Cauchy distributions share common quartiles, and the exponential and logistic share a common variance.

The second issue is the choice of design parameters, initial value and the set constant for the Delayed Robbins-Monro algorithm. The initial value is set each time at 0, the median for each distribution. The set constant is taken to be 3.6, approximately two standard deviations for the logistic $F(\{y-\mu\}/\sigma)$ with $\mu = 0$ and $\sigma = 1$. Although much advice exists for the optimal selection of each, all of it assumes that more is known about the response distribution than is commonly the case. The values selected should provide a reasonable opportunity for the strategy to prove its worth, but we leave the question of optimal design parameters for further study.

4.2.2 Analysis

In this section we present the methods of analysis used to characterize the performance of the proposed strategy. That strategy consists of both data collection and estimation procedures. Data collection is taken to be the Delayed Robbins-Monro algorithm acting on transformed responses. Estimation assumes a power logistic distribution for the transformed responses and is accomplished via maximum likelihood. The intent of this analysis is to establish whether jointly the proposed data collection and estimation procedures are feasible and supportive of the Chapter 3 results.

To this end--joint performance evaluation--we place the most emphasis on the quality of the observed estimates. Certainly, the estimates first need to fairly represent the actual quantiles sought. The measure of closeness adopted here is the root mean square error. We augment the information given by the $\sqrt{\text{mse}}$ with the examination of histograms of the empirical estimate distribution, outlier identification, and observed sample sizes. Each of the $\sqrt{\text{mse}}$ and supporting information help to answer joint performance questions.

A few questions must be answered with this data. How does the underlying response distribution influence the results? Does the procedure work well for all target quantiles? What role does sample size play? Does the ".75" quantile estimate hold any promise? Each question is considered in Section 4.4.

Although the final estimates are of primary importance, the data on which they are based must be informative. One analysis goal is to determine if the data collection approach involving the Delayed Robbins-Monro and transformed responses yields informative data relative to the estimation approach taken. The quality of the estimates lends some insight to this issue, but more exacting evaluation of the data collected can be achieved. From the results of Chapter 3 a desirable data collection strategy must play a supportive role to estimation. By that we mean that the design should be practically optimal and should promote conditions for estimate existence. How well DRM plays that role is examined.

4.3 Optimization

Maximum likelihood estimation of the extreme quantiles requires the use of optimization techniques. The quantile estimator (3.14) is an MLE by the invariance property of maximum likelihood if the estimators in that expression are themselves maximum likelihood estimators. Thus the immediate task, for quantile estimation, is to maximize the likelihood function with respect to the parameter vector $(\mu, \sigma, m)^T$. We choose equivalently to maximize the log-likelihood and begin by setting the expression in (3.14) equal to zero for each of the three parameters. The resulting system of equations, the solution for which is a stationary point, has no closed form solution and must be solved numerically. The optimization technique chosen for that task is the subject of this section.

Newton-like methods are commonly employed for problems of this type. Fletcher [1987] describes the Newton-Raphson procedure in a manner convenient for this discussion. We adopt that presentation here. The Newton-Raphson method begins by locally approximating the log-likelihood function l about the k^{th} parameter vector iterate $\theta^{(k)}$ with a second-order truncated Taylor series expansion. Denote the vector of first derivatives with respect to θ and evaluated at $\theta^{(k)}$ by $h^{(k)}$. Similarly denote the Hessian matrix evaluated at $\theta^{(k)}$ by $H^{(k)}$. Define $\delta = \theta - \theta^{(k)}$, and then write the expansion of $l(\theta)$ about $\theta^{(k)}$ as

$$l(\theta^{(k)} + \delta) \approx l(\theta^{(k)}) + h^{(k)^T}\delta + 1/2\,\delta^T H^{(k)}\delta. \tag{4.1}$$

The sum $\theta^{(k)} + \delta^{(k)}$ serves as the next iterate $\theta^{(k+1)}$, where $\delta^{(k)}$ maximizes the right hand side of (4.1). The solution for $\delta^{(k)}$ is given by solving $\delta^{(k)} = -H^{(k)^{-1}} \cdot h^{(k)}$ and is a maximum for the quadratic approximation providing that the Hessian matrix is negative definite. Consider that $\delta^{(k)}$ is interpretable as a correction to $\theta^{(k)}$ or as the directional step in the process. We use this interpretation later.

A variation on the above involves a substitution for the H. In our optimization problem, the Hessian matrix contains random variables and can be approximated with its mathematical expectation E(H). The principal advantage in doing this is mathematical tractability. The Newton-Raphson method with this substitution is termed the Method of Scores [Kendall and Stuart 1978]. The Method of Scores, with adjustments to follow, is the foundation for our optimization procedure.

It was necessary to adjust the Method of Scores because of the unacceptably low percentage of cases attaining convergence. Examination of many cases revealed that some data sets gave rise to poorly conditioned matrices corresponding to E(H). The specific problem was that E(H) sometimes approached negative semidefinite form. The inverse matrix then assumed values of large magnitude, forcing the directional step δ to be too large. Two adjustments were necessary to improve the situation. We determined that the power parameter m was contributing the most to the near indeterminacy when it was present. Therefore, rather than optimizing with respect to all three parameters simultaneously, we optimized in terms of μ and σ for each value of m along a suitable unidimensional grid. The best among the grid solutions found was taken to be the local optimum. When employed, this modification enhanced the procedures performance, but insufficiently. Thus a second modification was made. A restricted step rule was adopted, preventing a step greater than .1 for each parameter iterate [Fletcher 1987]. The intention is to slow the procedure for μ and σ until it nears a local maximum. The Method of Scores, with both adjustments, yielded reasonable estimates for our feasibility study.

4.4 Results

In this section we discuss the empirical results and relate them to the theoretical findings of Chapter 3. Feasibility issues are addressed first by focusing on specific procedural performance characteristics. Empirical support follows for the procedure with regard to optimal design and summary information.

4.4.1 Feasibility

To be considered feasible, the procedure must perform well with respect to its task, extreme quantile estimation. Additionally, it should possess some distinguishing *features to recommend* its use. Both points are discussed.

Bias of the target quantile estimates is reported in Table 4.2. In each cell, included are the true target quantile, the average of its estimates, the bias, and a 95% confidence interval for the bias using a normal approximation. Confidence intervals with an asterisk highlight those cases in which the bias cannot be considered significant at the .05 level. The bias appears slight for estimations in which the true response distribution form was logistic or exponential; where present, in nine of ten cases its nature was positive. For the Cauchy response form the bias varies with target quantile, appearing moderate, slight, and large for the .8, .9 and .95 quantiles respectively. The large negative bias for the .95 quantile indicates that either the power logistic fit was unable to accommodate the Cauchy's heavy tails, the design is unable to gather enough information near that quantile, or both. However, the estimate bias in general is not prohibitive and the procedure seems to perform reasonable well for, at least, logistic and exponential extreme quantile estimation.

The $\sqrt{\text{mse}}$ performance is recorded in Table 4.3 together with the number of data sets for which convergence was achieved. The procedure performs better for the logistic response distribution than for the exponential, and better for the exponential than for the Cauchy. That the logistic response distribution provides the best results is no surprise. Recall that the power logistic used to fit the transformed responses is the exact model when the true response distribution is logistic. The very poor performance for some Cauchy cases is due to the greater bias, in general, found in Cauchy quantile estimates. The magnitude of the $\sqrt{\text{mse}}$'s are encouraging for most

Table 4.2 Bias of PLTR Strategy

Sample Blocks	Response Distribution	Target Quantile (p) .8	.9	.95
	Cauchy	$x = 1.512$ $\bar{\hat{x}} = 1.770$ bias = .258 (.183, .333)	$x = 3.381$ $\bar{\hat{x}} = 3.297$ bias = -.084 (-.179, .011)*	$x = 6.936$ $\bar{\hat{x}} = 4.774$ bias = -2.162 (-2.311, -2.013)
15	Exponential	$x = 1.663$ $\bar{\hat{x}} = 1.789$ bias = .135 (.061, .209)	$x = 2.920$ $\bar{\hat{x}} = 3.061$ bias = .141 (.072, .210)	$x = 4.177$ $\bar{\hat{x}} = 4.091$ bias = -.086 (-.150, -.022)
	Logistic	$x = 1.386$ $\bar{\hat{x}} = 1.471$ bias = .085 (.035, .135)	$x = 2.197$ $\bar{\hat{x}} = 2.329$ bias = .132 (.084, .180)	$x = 2.944$ $\bar{\hat{x}} = 3.043$ bias = .099 (.055, .143)
	Cauchy	$x = 1.512$ $\bar{\hat{x}} = 1.765$ bias = .253 (.187, .319)	$x = 3.381$ $\bar{\hat{x}} = 3.311$ bias = -.070 (-.159, .019)*	$x = 6.936$ $\bar{\hat{x}} = 5.067$ bias = 1.869 (-2.026, -1.712)
20	Exponential	$x = 1.663$ $\bar{\hat{x}} = 1.792$ bias = .129 (.072, .186)	$x = 2.920$ $\bar{\hat{x}} = 3.074$ bias = .154 (.093, .215)	$x = 4.177$ $\bar{\hat{x}} = 4.209$ bias = .032 (-.030, .094)*
	Logistic	$x = 1.386$ $\bar{\hat{x}} = 1.435$ bias = .049 (-0, .098)*	$x = 2.197$ $\bar{\hat{x}} = 2.310$ bias = .113 (.072, .154)	$x = 2.944$ $\bar{\hat{x}} = 3.002$ bias = .058 (.023, .093)

NOTE: In this table x denotes x_{100p}

Table 4.3 Root-mean square error, standard error, and the number of convergences for the PLTR strategy

Sample Blocks	Response Distribution	Target Quantile .8	.9	.95
15	Cauchy	.854 .814 #456	.961 .957 #393	2.421 1.089 #206
	Exponential	.809 .798 #444	.758 .745 #453	.640 .634 #384
	Logistic	.557 .550 #460	.525 .508 #431	.471 .460 #423
20	Cauchy	.776 .734 #478	.906 .903 #393	2.196 1.153 #208
	Exponential	.643 .630 #470	.621 .602 #480	.605 .604 #410
	Logistic	.546 .544 #472	.466 .452 #470	.387 .383 #450

logistic and exponential cells of the design matrix. Express $\sqrt{mse}$'s in σ units, where σ is the population standard deviation. Since $\sigma = 1.81$ a $\sqrt{mse}$ of $.6 \approx \sigma/3$. The easier to estimate median has $\sqrt{mse}$'s of $\sigma/4$ to $\sigma/5$ for sample sizes consistent with those found in 15 and 20 blocks respectively [Bodt and Tingey 1987]. A limited comparison between this procedure and Wetherill's indicates that they may be comparable for estimation of the .8 logistic quantile.

Detailed comparison with other procedures is involved and beyond the scope of the feasibility study. Where procedures differ only in estimation, the data can be considered of near equal quality, and direct comparisons of performance can be made. We do this in the next section, comparing logistic and power logistic estimators. However, when the designs differ also, much attention must be given to fairly using initial inputs to the design for each procedure. For example, Wetherill's [1963] UDTR uses the Up and Down strategy, and his estimator is an average of stimulus levels generated by that sequential design. The stimulus levels selected depend on the initial design point and the spacing between levels; the optimum selection for each depends on the response distribution at hand. Similarly the DRM collects data in a manner dependent on the initial design point and spacing, and the quality of the power logistic estimate is a function of that data. Comparing these two fairly requires examination over a variety of input selections and will not be attempted here.

The convergence of the optimization procedure is a major consideration. Generally, convergence was achieved often for the limited number of blocks used. A 77% convergence rate was the minimum observed except when estimating the .95 Cauchy quantile where the percentage was a dismal 41%. Table 4.3 shows that with 20 blocks convergence is likely to occur more often than with 15 blocks. For example, for the .8 quantile, where convergence was not a problem regardless of distribution, 91% and 95% convergence is achieved for 15 blocks and 20 blocks respectively. This gain in convergence rate is not present in the Cauchy .9 and .95 quantile cases. The problem is always more severe for the .95 quantile than the .8 quantile; although, for the exponential response the convergence rate rises slightly going from the .8 to .9 quantile.

The actual sample size plays an interesting role in the formation of Table 4.3. Table 4.4 lists the average number of subjects required for each cell of the design matrix. Of interest is that for the logistic and exponential response distributions the $\sqrt{mse}$ in Table 4.3 is smaller, the further out in the tail one estimates. An initial reaction might be to glance at Table 4.4 and claim that the increased sample size is the obvious cause. In truth though, the power logistic only has available to it 15 or 20 blocks regardless of the quantile being estimated. The not so obvious reason for the smaller $\sqrt{mse}$'s is that in each case the variance of the transformed response distribution decreases over $n_c = 3, 7, 14$ – the strategies corresponding to the .8, .9 and .95 quantiles respectively. Then, as we expect, the populations with smaller variances give rise to estimates with smaller $\sqrt{mse}$'s. This reasoning does not apply to

Table 4.4 Subject requirements for PLTR strategy.

Sample Blocks	Response Distribution	Target Quantile .8	.9	.95
15	Cauchy	avg. 35.3 min. 26 max. 43	avg. 71.5 min. 44 max. 94	avg. 124.8 min. 81 max. 169
	Exponential	avg. 35.3 min. 25 max. 43	avg. 74.5 min. 49 max. 97	avg. 137.7 min. 88 max. 176
	Logistic	avg. 35.5 min. 26 max. 43	avg. 75.6 min. 44 max. 95	avg. 143.1 min. 82 max. 192
20	Cauchy	avg. 47.4 min. 35 max. 56	avg. 97.9 min. 71 max. 120	avg. 168.2 min. 127 max. 220
	Exponential	avg. 46.7 min. 33 max. 56	avg. 99.8 min. 73 max. 125	avg. 187.5 min. 133 max. 245
	Logistic	avg. 47.3 min. 35 max. 57	avg. 102.3 min. 71 max. 127	avg. 195.7 min. 109 max. 250

the Cauchy response distribution as its variance does not exist. Bias is contributing heavily to the larger $\sqrt{mse}$ for the .95 quantile estimate.

The ".75" quantile estimate was suggested in Section 4.2.1 as a way of using the power logistic model fit to estimate quantiles other than the largest quantile determined by the transformed response strategy. The ".75" power logistic quantile estimate proved to be biased in nearly all cells of the design matrix. Apparently the power logistic fit is limited to local smoothing about the .5 power logistic quantile. In that role it performs well since each of the .8, .9 and .95 quantile estimates considered here result from such smoothing. Further study is required to determine the size of the trust region.

The principal reason for choosing a 3-parameter model is greater flexibility in fitting the data. Robustness to distributional form is considered. In terms of estimate bias, only the Cauchy distribution proved to be a problem for the estimation technique employed. The results of the $\sqrt{mse}$ and convergence performance were similar, with the Cauchy distribution presenting most of the problem. Though our study is limited, the procedure does work for more than just the logistic response distribution.

4.4.2 Empirical Support

Section 3.2.1 determines that data can be collected optimally by placing all observations at the quantile to be estimated. In Section 3.2.2 we suggested that the DRM strategy represents a practical approach to this optimal data collection. We have also mentioned the possibility that summarizing the data through transformed responses may hinder estimation. In this section we discuss the empirical support for the DRM strategy and the claim that information is not being lost by choosing to estimate with transformed responses rather than the original response data.

The DRM strategy does collect data about the quantile to be estimated. Figure 4.1 shows a histogram of the final stimulus level tested for each of the 384 cases corresponding to the exponential response distribution, 15 blocks, and the .95 quantile. The assumption is that the last stimulus collected is representative of several data points gathered by the sequential strategy. The mean of those values is 4.112, very close to the true quantile value of 4.177. Similar histograms were formed for the other cells of the design matrix. Of note is that the bias observed in the estimators is reflected in the last stimulus histograms.

Another concern in selecting the design is that it yields data likely to satisfy the existence conditions given in Section 3.2.3. Figure 4.2 shows a typical data set gathered by the DRM strategy acting on transformed responses. Note that the conditions are satisfied.

Table 4.5 lists the $\sqrt{mse}$'s associated with power logistic and logistic estimation based on the exact same data from the design matrix. Our purpose here is to determine whether or not information loss results from using the transformed

INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
2.2	0	0	0.0	0.0
2.4	0	0	0.0	0.0
2.6	1	1	0.3	0.3
2.8	1	2	0.3	0.5
3	12	14	3.1	3.6
3.2	16	30	4.2	7.8
3.4	29	59	7.6	15.4
3.6	26	85	6.8	22.1
3.8	22	107	5.7	27.9
4	51	158	13.3	41.1
4.2	61	219	15.9	57.0
4.4	37	256	9.6	66.7
4.6	44	300	11.5	78.1
4.8	40	340	10.4	88.5
5	19	359	4.9	93.5
5.2	11	370	2.9	96.4
5.4	5	375	1.3	97.7
5.6	5	380	1.3	99.0
5.8	1	381	0.3	99.2
6	2	383	0.5	99.7
6.2	1	384	0.3	100.0
6.4	0	384	0.0	100.0
6.6	0	384	0.0	100.0

Figure 4.1 Histogram of final stimulus levels tested

Stimulus Levels	Original Responses (+ or -)	Transformed Responses (1 or 0)	Required Move
.00	-	0	up
1.80	-	0	up
3.60	+ + + + + + +	1	down
2.70	+ + + + + + +	1	down
2.10	+ -	0	up
2.55	+ + + + + + +	1	down
2.19	+ -	0	up
2.49	+ + + -	0	up
2.75	+ + + + + + +	1	down
2.52	+ + + + + + +	1	down
2.32	+ + + + + + +	1	down
2.14	-	0	up
2.31	+ + + + + + +	1	down
2.16	+ + + + + -	0	up
2.29	+ + -	0	up

Figure 4.2 Transformed logistic responses collected by the DRM strategy

Table 4.5 Root-mean square error comparison between power logistic and logistic estimation for cases in which convergence was achieved for both

Sample Blocks	Response Distribution	Target Quantile .8	.9	.95
15	Power Logistic	.547 #404	.534 #380	.475 #381
	Logistic	.724	.448	.439
20	Power Logistic	.552 #438	.465 #417	.385 #412
	Logistic	.804	.438	.354

response information only. For this purpose $\sqrt{\text{mse}}$ is the primary focus. The power logistic estimate has a smaller $\sqrt{\text{mse}}$ than the logistic for the .8 quantile. The reason is that with just an average of 35 or 47 observations for 15 blocks and 20 blocks, respectively, the logistic estimator delivers some values much larger than the power logistic estimates. For example, with 15 blocks the maximum power logistic estimate is 5.1 as compared to 8.6 for the logistic. That one value, 8.6, contributed approximately .1 to the logistic $\sqrt{\text{mse}}$. With more data for the .9 and .95 target quantiles, the maximum estimates are of comparable size. In Figures 4.3-4.4 histograms of both estimators are shown for the 20 blocks and the .9 quantile. As Table 4.5 suggests, little difference can be seen, though the logistic estimate is some what better. Similar observations are appropriate for the other direct comparisons of .9 and .95 quantile estimates indicating that if information loss is induced by using transformed responses, that it is slight.

5. DISCUSSION

The challenge of extreme quantile estimation in a binary response environment is immense. The most perplexing of the obstacles faced is the binary response itself. The very limited information that it holds relevant to a subject's individual tolerance makes the study of that random variable and its distribution difficult in even the most *narrowly focused estimation quests*, such as the pursuit of central tendency via the tolerance distribution median. Even given the extensive treatment the median estimation problem has received in the open literature, no generally accepted correct solution to the problem exists. But in extreme quantile estimation the challenge is greater. There the parametric assumptions made become more critical. Consider the difference between the .9 quantiles from a heavy versus light tailed distribution, and the ability of common goodness of fit measures to discern between such distributions. Without the distribution family at our disposal a priori, which among the ten to fifteen proposed for this environment should be used? Results from c-optimality considerations suggest that except for the most extreme quantiles, data should be gathered at the stimulus level corresponding to the target quantile. But without prior knowledge of a completely specified distribution that stimulus level is also unknown. Some sequential methods overcome this hurdle by converging in probability to the desired stimulus level, but also they exhibit bias for finite samples. Among the many sequential strategies available, which is best suited for extreme quantile estimation? Finally, the most common estimation, maximum likelihood, is often hampered by data sets which fail to produce unique estimates and likelihood equations with no closed form solution for the stationary values. It is therefore with open eyes that one must proceed when engaging this challenge, and we have attempted to do so here.

To a great extent this work joins established concepts and focuses them on extreme quantile inference. We believe that the approach proposed here, Power Logistic Transformed Response strategy (PLTR), represents a contribution in

```
INTERVAL                                                                                       FREQUENCY PERCENTAGE
NAME           5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80  INT. CUM. INT.  CUM.
          +----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+
 1        +                                                                                  0    0   0.0   0.0
 1.2      +                                                                                  0    0   0.0   0.0
 1.4      +XX                                                                                2    2   0.5   0.5
 1.6      +XXXXXXXXXXX                                                                      11   13   2.6   3.1
 1.8      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                                        41   54   9.8  12.9
 2        +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                   62  116  14.9  27.8
 2.2      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX* 103  219  24.7  52.5
 2.4      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   78  297  18.7  71.2
 2.6      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                                     44  341  10.6  81.8
 2.8      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                                              35  376   8.4  90.2
 3        +XXXXXXXXXXXXXXXXXXXXXXXXXX                                                       26  402   6.2  96.4
 3.2      +XXXXXX                                                                            6  408   1.4  97.8
 3.4      +XXX                                                                               3  411   0.7  98.6
 3.6      +XX                                                                                2  413   0.5  99.0
 3.8      +X                                                                                 1  414   0.2  99.3
 4        +X                                                                                 1  415   0.2  99.5
 4.2      +                                                                                  0  415   0.0  99.5
 4.4      +                                                                                  0  415   0.0  99.5
 4.6      +                                                                                  0  415   0.0  99.5
 4.8      +                                                                                  0  415   0.0  99.5
 5        +XX                                                                                2  417   0.5 100.0
          +----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+
               5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80
```

Figure 4.3 Empirical estimate density resulting from logistic estimation on original responses

```
INTERVAL                                                                                          FREQUENCY PERCENTAGE
NAME          5    10   15   20   25   30   35   40   45   50   55   60   65   70   75   80     INT. CUM. INT.  CUM.
        +----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+
 1      +                                                                                        0    0   0.0   0.0
 1.2    +                                                                                        0    0   0.0   0.0
 1.4    +XX                                                                                      2    2   0.5   0.5
 1.6    +XXXXXXXXXXXX                                                                           12   14   2.9   3.4
 1.8    +XXXXXXXXXXXXXXXXXXXXXXXXXXX                                                            27   41   6.5   9.8
 2      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                        63  104  15.1  24.9
 2.2    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*      81  185  19.4  44.4
 2.4    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*      86  271  20.6  65.0
 2.6    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                              57  328  13.7  78.7
 2.8    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                                             42  370  10.1  88.7
 3      +XXXXXXXXXXXXXXXXXXXXXXXX                                                               24  394   5.8  94.5
 3.2    +XXXXXXXXXXXXX                                                                          13  407   3.1  97.6
 3.4    +X                                                                                       1  408   0.2  97.8
 3.6    +XX                                                                                      2  410   0.5  98.3
 3.8    +                                                                                        0  410   0.0  98.3
 4      +XX                                                                                      2  412   0.5  98.8
 4.2    +XXX                                                                                     3  415   0.7  99.5
 4.4    +X                                                                                       1  416   0.2  99.8
 4.6    +                                                                                        0  416   0.0  99.8
 4.8    +X                                                                                       1  417   0.2 100.0
 5      +                                                                                        0  417   0.0 100.0
        +----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+
              5    10   15   20   25   30   35   40   45   50   55   60   65   70   75   80
```

Figure 4.4 Empirical estimate density resulting from power logistic estimation on transformed responses addressing many of the difficulties discussed, but we recognize the potential for its improvement. The PLTR begins with a transformation of responses as given in Table 2.1. Its principal advantage is that it reduces the problem to one of median estimation, thereby allowing us the luxury of open literature information on that problem. Moreover, we avoid the bias usually associated with sequential strategies operating in the distribution tail. The transformation actually summarizes the original response information, and the empirical evidence suggests that not a great deal of information is lost through condensation of the sample space. This transformation is superior to just fixing multiple samples for each stimulus because it permits a sequential strategy acting on it to abandon stimulus levels early when a failure is encountered, a more frugal use of resources. Another summarization procedure [McLeish and Tosh 1983] shows promise, but it depends to some, possibly limited, extent on the logistic distribution whereas transformed responses do not. The independence of distributional form assists us in collecting data for a c-optimal design for a variety of distribution families. Additionally, even if estimation is to be made using the original responses, the transformed response method can assist the experimenter in gathering data local to the extreme quantiles. Our conclusion is that the use of the transformed responses with a good sequential strategy is highly recommended for data collection. More research is required to assert that estimation with respect to the transformed response distribution is the the best way to go, though our initial results are encouraging.

Under transformed responses the target quantile is local to the median of the transformed response distribution. With the problem reduced to median estimation we chose the Stochastic Approximation Method of Robbins and Monro [1951] as our sequential strategy. Asymptotically the procedure is consistent and therefore desirable with respect to c-optimality design requirements when the power logistic parameter m is known. This claim also holds for several parametric assumptions besides the power logistic model employed here. Further, the small sample performance of the strategy in Chapter 4 as a design and in Bodt and Tingey [1986] as a design used with maximum likelihood estimation has been good. Finally, we showed in Section 3.2.2 that data satisfying *II* from Theorem 3.1 will likely result from Robbins-Monro implementation, thereby ensuring MLE existence for μ and σ when m is known.

Certainly, other design possibilities exist. Wetherill et al. [1966] used for this purpose Dixon and Mood's [1948] Up and Down Method, then considered a good sequential strategy. It calls for equal spacings between adjacent stimulus levels making it conducive for use with his estimator w, a nonparametric estimator discussed in Chapter 2. Direct comparison between UDTR and PLTR is difficult because the stimulus levels tested are different, but in a rough comparison we saw no real difference in final results for the logistic distribution. To our knowledge his procedure has not been examined for other response forms. Einbinder [1973] substituted Langlie's design for the Up and Down and a 3-parameter Weibull MLE on the original responses for $\overline{w}$. No extensive check of its performance has been done. Other designs might be suitable, such as Wu's [1985] EQRC, but in our study [Bodt and Tingey 1986] we found no significant advantage to using it over the Delayed Robbins-Monro with maximum likelihood estimation after the data had been collected. Thus, we have selected one of the best possible strategies for the sequential design task.

Many distributional forms could have been used to estimate the target quantile, but we chose the power logistic. It possesses two attractive qualities. The first is that it has great intuitive appeal if the true underlying response distribution is logistic. Then the power logistic is the exact distribution corresponding to the transformed response. This enabled us to compare the estimation of extreme quantiles using original responses and a logistic distribution to transformed responses and the power logistic distribution. The second is that the third parameter provides greater flexibility in modeling the distribution in the usual case where it is unknown.

One problem with our selection is the difficulty associated with solving the likelihood equations. We established conditions under which existence is guaranteed, and we tailored our grid optimization approach to those results. Specifically, for each fixed value of one parameter, conditions were established under which the MLEs for the other two parameters exist uniquely. The grid approach combined with a restricted step Method of Scores allowed us to select the best among the two parameter maxima computed. It is possible that a reparameterization is called for, or possibly an alternate parametric assumption.

The choice of the power logistic was important for our stated purposes, but other assumptions may be better suited to this problem. Forms such as the cubic logistic [Morgan 1985] and 3-parameter Weibull may prove easier to implement with regard to maximum likelihood estimation. Possibly some of the distributions suggested in Chapter 2 as robust procedures could be used. However, when choosing among them, we must recognize that their tail behavior is not as important as their flexibility about the median if used with transformed responses.

In closing, we have demonstrated a feasible new approach for extreme quantile estimation in binary response models. Supporting PLTR are encouraging Monte Carlo results and analytical findings suggesting the appropriateness of its component parts. However, much more can be done to extend this work. We feel that collecting data in the manner described is probably a good start but parametric assumptions are an open question. The analytical results could be extended for these other distributions providing a more general treatment. Comprehensive performance studies similar to that of Bodt and Tingey [1986] need to be carried out before suggesting its general use. For example, the influence of varying initial design points, DRM constant settings, and target quantiles on the estimates should be examined over a wide range of true response functions. Finally, direct comparison with other techniques such as the Alexander Extreme Value Design, the work of McLeish and Tosh [1983] and Wetherill's UDTR strategy is needed.

INTENTIONALLY LEFT BLANK.

BIBLIOGRAPHY

Anbar, D. [1978], "A Stochastic Newton-Raphson Method," Journal of Statistical Planning and Inference, Vol. 2, pp. 153-163.

Aranda-Urdaz, F.J. [1981], "On Two Families of Transformations to Additivity for Binary Response Data," Biometrika, Vol. 68, pp. 357-364.

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., Silverman, E. [1955], "An Empirical Distribution Function for Sampling With Incomplete Information," Annals of Mathematical Statistics, Vol. 26, pp. 641-647.

Baker, R.J., Nelder, J.A. [1978], The GLIM System, Oxford: Clarendon.

Bartlett, M.S. [1946], "A Modified Probit Technique for Small Probabilities," Supplement Journal of the Royal Statistical Society, Vol. 8, pp. 113-117.

Berkson, J. [1944], "Application of the Logistic Function to Bioassay," Journal of the American Statistical Association, Vol. 39, pp. 357-365.

Bliss, C.I. [1934a], "The Method of Probits," Science, Vol. 79, pp. 38-39.

Bliss, C.I. [1934b], "The Method of Probits--a Correction," Science, Vol. 79, pp. 409-410.

Bodt, B.A., Tingey, H.B. [1986], "Considerations in Small Sample Quantal Response Testing," Proceedings of the Thirty-First Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 86-2, US Army Research Office, pp. 65-83.

Bodt, B.A., Tingey, H.B. [1987], "A Comparison of Two Sensitivity Testing Procedures with Implications for Sample Size Determination," Proceedings of the Thirty-Second Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 87-2, US Army Research Office, pp. 293-316.

Brown, B.W. [1967], "Quantal Response Assays," Statistics in Endocrinology, J.W. McArthur and T. Colton, eds., Cambridge: The MIT Press.

Brownlee, K.A., Hodges, J.L. Jr., Rosenblatt, M. [1953], "The Up-and-Down Method with Small Samples," Journal of the American Statistical Association, Vol. 48, pp. 262-277.

Chernoff, H. [1962], "Optimal Design of Experiments," Technical Report No. 82, Applied Mathematics and Statistics Laboratories, Stanford University.

Chernoff, H. [1979], Sequential Analysis and Optimal Design, Revised Edition, Philadelphia: SIAM.

Chung, K.L. [1954], "On a Stochastic Approximation Method," Annals of Mathematical Statistics, Vol. 25, pp. 463-483.

Cochran, W.G., Davis, M. [1964], "Stochastic Approximation to the Median Effective Dose in Bioassay," Stochastic Models in Medicine and Biology, J. Garland ed., pp. 281-300, Madison: University of Wisconsin Press.

Copenhaver, T.W., Mielke, P.W. [1977], "Quantit Analysis: a Quantal Assay Refinement," Biometrics, Vol. 33, pp. 175-186.

Davis, M. [1971], "Comparison of Sequential Bioassays in Small Samples," Journal of the Royal Statistical Society, Ser. B, Vol. 33, pp. 78-87.

DiDonato, A.R., Jarnagin, M.P. Jr. [1972], "Use of the Maximum Likelihood Method Under Quantal Responses for Estimating the Parameters of a Normal Distribution and its Application to an Armor Penetration Problem," Technical Report TR-2846, Naval Weapons Laboratory.

Dixon, W.J., Mood, H.M. [1948], "A Method for Obtaining and Analyzing Sensitivity Data," Journal of the American Statistical Association, Vol. 43, pp. 109-126.

Dvoretzky, A. [1956], "On Stochastic Approximation," Proceedings of the Third Berkeley Symposium, University of Californial, Berkeley, pp. 39-55.

Egger, M.J. [1979], "Power Transformations to Achieve Symmetry in Quantal Bioassay," Technical Report 47, Division of Biostatistics, Stanford University.

Einbinder, S.K. [1973], "Reliability Models and Estimation of Stress-Strength," Dissertation, Polytechnic Institute of Brooklyn.

Fedorov, V.V. [1972], Theory of Optimal Experiments, New York: Academic Press.

Finney, D.J. [1978], Statistical Method in Biological Assay, London: Charles Griffen & Company LTD.

Fletcher, R. [1987], Practical Methods of Optimization, New York: John Wiley & Sons LTD.

Ford, I., Tornsey, B., Wu, C.F.J. [1988], "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Nonlinear Problems," Journal of the Royal Statistical Society, Ser. B., preprint.

Guerrero, V.M., Johnson, R.A. [1982], "Use of the Box-Cox Transformation with Binary Response Models," Biometrika, Vol. 69, pp. 309-314.

Hamilton, M.A. [1979], "Robust Estimates of the ED50," Journal of the American Statistical Association, Vol. 74, pp. 344-354.

Hoel, P.G., Jennrich, R.I. [1979], "Optimal Designs for Dose Response Experiments in Cancer Research," Biometrika, Vol. 66, pp. 307-316.

Hsi, B.P. [1969], "The Multiple Sample Up-and-Down Method in Bioassay," Journal of the American Statistical Association, Vol. 64, pp. 147-162.

Kärber, G. [1931], "Beitrag zur Kollektiven Behandlung Pharmalcologischer Reihenversuche," Arch. Exp. Path. Pharmak., Vol. 162, pp. 480-487.

Kendall, M., Stuart, A. [1979], The Advanced Theory of Statistics, Vol. 2, Ed. 4, New York: Macmillan Publishing Co., Inc.

Kesten, H. [1958], "Accelerated Stochastic Approximation," Annals of Mathematical Statistics, Vol. 29, pp. 41-59.

Lai, T.L., Robbins, H. [1979], "Adaptive Design and Stochastic Approximation," Annals of Statistics, Vol. 7, pp. 1196-1221.

Langlie, H.J. [1962], "A Reliability Test Method for 'One-Shot' Items," Aeronutronic Publication No. U-1792.

Little, R.E. [1974], "A Mean Square Error Comparison of Certain Median Response Estimates for the Up-and-Down Method with Small Samples," Journal of the American Statistical Association, Vol. 69, pp. 202-206.

Little, R.E. [1976], "Planning Quantal Response Tests for Ordnance Devices: The Two Point Strategy," Proceedings of the Twenty-First Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 76-2, US Army Research Office, pp. 151-165.

McLeish, D., Tosh, D. [1983], "The Estimation of Extreme Quantiles in Logit Bioassay," Biometrika, Vol. 70, pp. 625-632.

Mielke, P.W. [1972], "Asymptotic Behavior of Two-Sample Tests Based on Powers of Ranks for Detecting Scale and Location Alternatives," Journal of the American Statistical Association, Vol. 67, pp. 850-854.

Morgan, B.J.T. [1985], "The Cubic Logistic Model for Quantal Assay Data," Applied Statistics, Vol. 34, pp. 105-113.

Prentice, R.L. [1976], "A Generalization of Probit and Logit Methods for Dose Response Curves," Biometrics, Vol. 32, pp. 761-768.

Robbins, H., Monro, S. [1951], "A Stochastic Approximation Method," Annals of Mathematical Statistics, Vol. 22, pp. 400-407.

Rothman, D. [1959], "Sequential Design of Multi-Level Pulse Testing of Rocket Engines when Each Test Series Must Terminate at the First Level which Gives Instability," Research Report 59-28, Rocketdyne.

Rothman, D., Alexander, M.J., Zimmerman, J.M. [1965], "The Design and Analysis of Sensitivity Experiments," NASA CR-62026, Vol. 1.

Rubinstein, R.Y. [1981], Simulation and the Monte Carlo Method, New York, John Wiley & Sons Inc..

Silvapulle, M.J. [1981], "On the Existence of Maximum Likelihood Estimators for the Binomial Response Model," Journal of the Royal Statistical Society, Vol. 43, pp. 310-313.

Silvey, D. [1980], Optimal Design, London: Chapman and Hall.

Trevan, J.W. [1927], "The Error of Determination of Toxicity," Proc. R. Soc., Vol. B101, pp. 483-514.

Wedderburn, R.W.M. [1976], "On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models," Biometrika, Vol. 63, pp. 27-32.

Wetherill, G.B. [1963], "Sequential Estimation of Quantal Response Curves," Journal of the Royal Statistical Society, Vol. 25, pp. 1-48.

Wetherill, G.B., Chen, H., Vasudeva, R.B. [1966], "Sequential Estimation of Quantal Response Curves: A New Method of Estimation," Biometrika, Vol. 53, pp. 439-454.

Wetherill, G.B. [1976], Sequential Methods in Statistics, London: Chapman and Hall.

Wu, C.F.J. [1985], "Efficient Sequential Designs with Binary Data," Journal of the American Statistical Association, Vol. 8, pp. 974-984.

Wu, C.F.J. [1988], "Optimal Design for Percentile Estimation of a Quantal Response Curve," Proceedings of the First International Conference-Workshop on Optimal Design and Analysis of Experiments, Neochatel, Switzerland.

No of Copies	Organization
1	Office of the Secretary of Defense OUSD(A) Director, Live Fire Testing ATTN: James F. O'Bryon Washington, DC 20301-3110
ıclass., unlimited) 12 ıclass., limited) 2 assified) 2	Administrator Defense Technical Info Center ATTN: DTIC-DDA Cameron Station Alexandria, VA 22304-6145
1	HQDA (SARD-TR) WASH DC 20310-0001
1	Commander US Army Materiel Command ATTN: AMCDRA-ST 5001 Eisenhower Avenue Alexandria, VA 22333-0001
1	Commander US Army Laboratory Command ATTN: AMSLC-DL Adelphi, MD 20783-1145
2	Commander Armament RD&E Center US Army AMCCOM ATTN: SMCAR-MSI Picatinny Arsenal, NJ 07806-5000
2	Commander Armament RD&E Center US Army AMCCOM ATTN: SMCAR-TDC Picatinny Arsenal, NJ 07806-5000
1	Director Benet Weapons Laboratory Armament RD&E Center US Army AMCCOM ATTN: SMCAR-CCB-TL Watervliet, NY 12189-4050
1	Commander US Army Armament, Munitions and Chemical Command ATTN: SMCAR-ESP-L Rock Island, IL 61299-5000
1	Commander US Army Aviation Systems Command ATTN: AMSAV-DACL 4300 Goodfellow Blvd. St. Louis, MO 63120-1798

No of Copies	Organization
1	Director US Army Aviation Research and Technology Activity Ames Research Center Moffett Field, CA 94035-1099
1	Commander US Army Missile Command ATTN: AMSMI-RD-CS-R (DOC) Redstone Arsenal, AL 35898-5010
1	Commander US Army Tank-Automotive Command ATTN: AMSTA-TSL (Technical Library) Warren, MI 48397-5000
1	Director US Army TRADOC Analysis Command ATTN: ATAA-SL White Sands Missile Range, NM 88002-5502
(Class. only) 1	Commandant US Army Infantry School ATTN: ATSH-CD (Security Mgr.) Fort Benning, GA 31905-5660
(Unclass. only) 1	Commandant US Army Infantry School ATTN: ATSH-CD-CSO-OR Fort Benning, GA 31905-5660
1	Air Force Armament Laboratory ATTN: AFATL/DLODL Eglin AFB, FL 32542-5000

<u>Aberdeen Proving Ground</u>

No of Copies	Organization
2	Dir, USAMSAA ATTN: AMXSY-D AMXSY-MP, H. Cohen
1	Cdr, USATECOM ATTN: AMSTE-TO-F
3	Cdr, CRDEC, AMCCOM ATTN: SMCCR-RSP-A SMCCR-MU SMCCR-MSI
1	Dir, VLAMO ATTN: AMSLC-VL-D

No of Copies	Organization
1	Commander US Army Materiel Command ATTN: AMCAE, C. Chapin 5001 Eisenhower Avenue Alexandria, VA 22333-0001
1	Professor Henry B. Tingey Ewing Hall Department of Mathematical Sciences University of Delaware Newark, DE 19716-5000
	Aberdeen Proving Ground
1	Cdr, USACSTA ATTN: STECS-DA-PS, V. Visnaw

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. BRL Report Number BRL-TR-3089 Date of Report MAR 90

2. Date Report Received ______

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) ______

4. Specifically, how is the report being used? (Information source, design data, procedure, source of ideas, etc.) ______

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided, or efficiencies achieved, etc? If so, please elaborate. ______

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.) ______

CURRENT ADDRESS

Name ______

Organization ______

Address ______

City, State, Zip Code ______

7. If indicating a Change of Address or Address Correction, please provide the New or Correct Address in Block 6 above and the Old or Incorrect address below.

OLD ADDRESS

Name ______

Organization ______

Address ______

City, State, Zip Code ______

(Remove this sheet, fold as indicated, staple or tape closed, and mail.)

---FOLD HERE---

DEPARTMENT OF THE ARMY

Director
U.S. Army Ballistic Research Laboratory
ATTN: SLCBR-DD-T
Aberdeen Proving Ground, MD 210[illegible]-5066
OFFICIAL BUSINESS

NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS PERMIT No 0001, APG, MD

POSTAGE WILL BE PAID BY ADDRESSEE

Director
U.S. Army Ballistic Research Laboratory
ATTN: SLCBR-DD-T
Aberdeen Proving Ground, MD 21005-9989

---FOLD HERE---